Equazioni Differenziali Stocastiche e Deriva Linguistica: Brownian Motion nei Cambiamenti Fonetici

Un'Analisi Matematica Rigorosa dell'Evoluzione Fonologica attraverso Processi Stocastici Continui

---

Motivazioni Epistemologiche

1. Introduzione: La Convergenza tra Matematica Stocastica e Linguistica Diacronica

La linguistica storica ha tradizionalmente affrontato i fenomeni di cambiamento linguistico attraverso metodologie prevalentemente qualitative, basate sulla comparazione filologica e sulla ricostruzione etimologica. Tuttavia, l'emergere della linguistica quantitativa e computazionale nel corso del ventesimo secolo ha aperto prospettive inedite per la modellizzazione matematica dei processi evolutivi del linguaggio. In particolare, la teoria dei processi stocastici, sviluppata originariamente per descrivere fenomeni fisici come il moto browniano delle particelle e successivamente applicata con straordinario successo alla genetica delle popolazioni, offre un quadro concettuale e analitico straordinariamente potente per comprendere la dinamica dei cambiamenti fonetici, morfologici e lessicali.

L'idea centrale che anima questo approccio è quella di concepire l'evoluzione linguistica non come una successione deterministica di trasformazioni governate da leggi fonetiche immutabili, bensì come un processo intrinsecamente probabilistico in cui fluttuazioni casuali, pressioni selettive e contingenze storiche interagiscono in modo complesso. Questa prospettiva rappresenta un superamento della visione neogrammatica classica, secondo cui le leggi fonetiche opererebbero in modo meccanico e senza eccezioni, e si avvicina maggiormente alle moderne concezioni dell'evoluzione biologica, dove la stocasticità gioca un ruolo fondamentale accanto alla selezione naturale.

La modellizzazione attraverso equazioni differenziali stocastiche permette di catturare formalmente la tensione tra regolarità e variabilità che caratterizza tutti i fenomeni linguistici. Da un lato, osserviamo tendenze direzionali nel cambiamento linguistico, come la Grande Mutazione Vocalica inglese o la lenizione consonantica nelle lingue romanze, che sembrano suggerire l'esistenza di forze sistematiche. Dall'altro, la variabilità inter-individuale, la stratificazione sociale, le influenze di contatto e numerosi altri fattori introducono elementi di casualità che rendono l'esito finale di questi processi fondamentalmente imprevedibile nel dettaglio, pur essendo caratterizzabile statisticamente.

2. Il Moto Browniano come Paradigma Fondamentale

Il moto browniano, osservato per la prima volta dal botanico Robert Brown nel 1827 studiando il movimento erratico di granuli di polline sospesi in acqua, e successivamente spiegato teoricamente da Albert Einstein nel suo celebre lavoro del 1905, costituisce il prototipo dei processi stocastici continui. La teoria matematica sviluppata da Norbert Wiener negli anni Venti del Novecento ha fornito una rigorosa fondazione alla nozione di processo di Wiener, che rappresenta la formalizzazione matematica del moto browniano.

Un processo di Wiener $W(t)$ è caratterizzato dalle seguenti proprietà fondamentali: primo, $W(0) = 0$ quasi certamente; secondo, il processo ha incrementi indipendenti, il che significa che per ogni sequenza di tempi $t_0 < t_1 < ... < t_n$, le variabili aleatorie $W(t_1) - W(t_0)$, $W(t_2) - W(t_1)$, ..., $W(t_n) - W(t_{n-1})$ sono statisticamente indipendenti; terzo, ogni incremento $W(t+s) - W(s)$ è distribuito normalmente con media zero e varianza t; quarto, le traiettorie sono continue ma non differenziabili quasi ovunque.

Queste proprietà matematiche possiedono profonde implicazioni quando le traduciamo nel contesto linguistico. L'indipendenza degli incrementi corrisponde all'assunzione che i cambiamenti fonetici in intervalli temporali disgiunti siano sostanzialmente indipendenti, una volta condizionati sullo stato presente del sistema. La distribuzione normale degli incrementi riflette il teorema del limite centrale applicato alla somma di numerosi piccoli effetti casuali individuali. La continuità ma non differenziabilità delle traiettorie cattura l'idea che il cambiamento linguistico sia graduale ma caratterizzato da fluttuazioni continue a tutte le scale temporali.

3. Spazi Fonetici e Geometria del Cambiamento Sonoro

Per applicare la teoria dei processi stocastici ai fenomeni linguistici, dobbiamo innanzitutto costruire una rappresentazione geometrica appropriata dello spazio fonetico. La fonetica articolatoria e acustica forniscono coordinate naturali per questa parametrizzazione. Nel caso delle vocali, lo spazio può essere rappresentato efficacemente attraverso le prime due formanti, F_1 e F_2, che corrispondono rispettivamente al grado di apertura e al grado di anteriorità della vocale. Questo approccio, sviluppato sistematicamente da Gunnar Fant e altri pionieri della fonetica acustica, permette di mappare ogni vocale in un punto del piano F_1-F_2.

Più precisamente, consideriamo uno spazio fonetico astratto X, che possiamo assumere essere un sottoinsieme aperto e limitato di $\mathbb{R}^n$, dove n rappresenta il numero di parametri fonetici rilevanti. Per le vocali, potremmo avere $n = 2$ (usando F_1 e F_2) oppure $n = 3$ (includendo anche F_3 o la durata). Per le consonanti, lo spazio parametrico diventa necessariamente più complesso, richiedendo dimensioni per il modo di articolazione, il luogo di articolazione, la sonorità, l'aspirazione, e così via.

La distanza tra due suoni nello spazio X può essere definita utilizzando metriche appropriate. La distanza euclidea standard rappresenta una prima approssimazione, ma ricerche più sofisticate hanno proposto metriche che tengono conto della percezione umana, come la distanza di Bark per le frequenze formantiche. Formalmente, possiamo scrivere una

metrica perceptualmente motivata come $d(x, y) = \sqrt{[\Sigma_i w_i(x_i - y_i)^2]}$, dove i pesi w_i riflettono la salienza percettiva delle diverse dimensioni.

4. Equazioni Differenziali Stocastiche: Il Formalismo Generale

Un'equazione differenziale stocastica rappresenta l'estensione stocastica di un'equazione differenziale ordinaria, incorporando termini di rumore che catturano le fluttuazioni casuali. La forma generale di una SDE (Stochastic Differential Equation) è:

$$dX(t) = \mu(X(t), t)\, dt + \sigma(X(t), t)\, dW(t)$$

dove $X(t) \in \mathbb{R}^n$ rappresenta lo stato del sistema al tempo t, $\mu: \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ è il coefficiente di drift (deriva), $\sigma: \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^{n \times m}$ è il coefficiente di diffusione, e W(t) è un processo di Wiener m-dimensionale.

Nel contesto linguistico, X(t) rappresenta la posizione media di un fono particolare nello spazio fonetico al tempo t. Il termine di drift μ cattura le tendenze sistematiche nel cambiamento fonetico, come le attrazioni verso configurazioni articolatorie più economiche o gli effetti di coarticolazione sistematici. Il termine di diffusione σ dW(t) rappresenta le fluttuazioni casuali dovute alla variabilità inter-individuale, agli errori di percezione e produzione, alle influenze sociolinguistiche e a numerosi altri fattori stocastici.

La distinzione tra drift e diffusione è fondamentale. Il drift rappresenta la componente deterministica del cambiamento, quella che potremmo osservare facendo la media su un numero infinito di realizzazioni indipendenti del processo. La diffusione rappresenta invece la dispersione attorno a questa tendenza media, quantificando l'incertezza intrinseca del processo evolutivo.

5. L'Equazione di Fokker-Planck e le Distribuzioni di Probabilità

Mentre un'equazione differenziale stocastica descrive l'evoluzione di singole traiettorie, spesso siamo interessati non tanto alla traiettoria particolare seguita da un suono specifico, quanto alla distribuzione di probabilità dell'intero ensemble di parlanti in una popolazione. Questo passaggio dalla prospettiva lagrangiana (che segue particelle individuali) alla prospettiva euleriana (che descrive campi di densità) è mediato dall'equazione di Fokker-Planck.

Se X(t) soddisfa la SDE $dX(t) = \mu(X, t)\, dt + \sigma(X, t)\, dW(t)$, allora la densità di probabilità $p(x, t)$ dell'essere nello stato x al tempo t evolve secondo l'equazione di Fokker-Planck:

$$\partial p/\partial t = -\nabla \cdot [\mu(x, t)\, p(x, t)] + (1/2)\, \nabla \cdot \nabla \cdot [D(x, t)\, p(x, t)]$$

dove $D(x, t) = \sigma(x, t)\, \sigma(x, t)^T$ è il tensore di diffusione, e ∇ indica l'operatore gradiente mentre $\nabla \cdot$ denota la divergenza.

Questa equazione rappresenta un bilancio di probabilità: il primo termine sul lato destro descrive il trasporto di probabilità dovuto al drift, mentre il secondo termine descrive la dispersione dovuta alla diffusione. È un'equazione deterministica per una densità di

probabilità, e rappresenta la controparte probabilistica delle equazioni di continuità della fisica dei fluidi.

Nel contesto linguistico, p(x, t) rappresenta la frazione della popolazione di parlanti che al tempo t produce un particolare fono nella regione x dello spazio fonetico. L'evoluzione di questa distribuzione cattura come la comunità linguistica nel suo complesso si muove attraverso lo spazio fonetico, pur mantenendo una certa dispersione attorno al valore modale.

6. Il Teorema di Itô e l'Integrazione Stocastica

Un aspetto tecnicamente cruciale nella teoria delle equazioni differenziali stocastiche è la definizione rigorosa dell'integrale stocastico $\int \sigma(X(s), s)\, dW(s)$. A differenza degli integrali ordinari, gli integrali stocastici presentano sottilità matematiche profonde dovute al fatto che le traiettorie del processo di Wiener non sono di variazione limitata.

Kiyoshi Itô ha risolto questo problema negli anni Quaranta sviluppando una teoria dell'integrazione stocastica basata su limiti in L^2. Il calcolo di Itô introduce una regola di derivazione modificata, nota come lemma di Itô, che rappresenta l'analogo stocastico della regola della catena. Per una funzione sufficientemente regolare $f: \mathbb{R}^n \to \mathbb{R}$ e un processo $X(t)$ che soddisfa $dX(t) = \mu(X, t)\, dt + \sigma(X, t)\, dW(t)$, il lemma di Itô afferma che:

$$df(X(t)) = [\partial f/\partial t + \mu^T \nabla f + (1/2)\, \mathrm{Tr}(\sigma\sigma^T Hf)]\, dt + (\nabla f)^T \sigma\, dW(t)$$

dove Hf denota la matrice hessiana di f e Tr indica la traccia. Il termine aggiuntivo $(1/2)\, \mathrm{Tr}(\sigma\sigma^T Hf)$, assente nel calcolo ordinario, riflette la variazione quadratica non nulla del processo di Wiener.

Questa formula ha importanti conseguenze per la modellizzazione linguistica. Per esempio, se vogliamo studiare come evolve non la posizione fonetica stessa ma qualche funzione di essa (come la distanza da una configurazione di riferimento o l'energia articolatoria), dobbiamo applicare il lemma di Itô. Il termine di secondo ordine implica che anche in assenza di drift nella variabile originale, funzioni non lineari di quella variabile possono sviluppare un drift effettivo dovuto esclusivamente alla diffusione stocastica.

7. Modelli con Drift e Selezione: Equazioni di Ornstein-Uhlenbeck

Un caso particolare estremamente rilevante è rappresentato dal processo di Ornstein-Uhlenbeck, che incorpora un meccanismo di ritorno verso la media. L'equazione differenziale stocastica è:

$$dX(t) = -\theta\, [X(t) - \mu]\, dt + \sigma\, dW(t)$$

dove $\theta > 0$ è il coefficiente di ritorno verso la media, μ è il livello di equilibrio, e σ è la volatilità.

Questo modello è particolarmente appropriato per situazioni linguistiche in cui esistono configurazioni fonetiche preferenziali verso cui il sistema tende a convergere, ma da cui fluttuazioni casuali lo fanno continuamente deviare. Il parametro θ quantifica la forza della

pressione di selezione: valori grandi di θ corrispondono a una forte tendenza restaurativa, mentre valori piccoli permettono escursioni più ampie.

La soluzione esplicita del processo di Ornstein-Uhlenbeck può essere scritta come:

$$X(t) = X(0)\, e^{-\theta t} + \mu\,(1 - e^{-\theta t}) + \sigma \int_0^t e^{-\theta(t-s)}\, dW(s)$$

Questa formula mostra che la soluzione consiste in tre componenti: un decadimento esponenziale delle condizioni iniziali, una convergenza esponenziale verso il valore medio μ, e un termine integrale stocastico che rappresenta l'effetto cumulativo delle fluttuazioni.

La distribuzione stazionaria, se esiste, è normale con media μ e varianza $\sigma^2/(2\theta)$. Questo risultato implica che nel lungo termine, se non ci sono altre forze in gioco, la distribuzione fonetica si stabilizza attorno a una configurazione di equilibrio con una dispersione determinata dal rapporto tra la forza delle fluttuazioni e la forza della selezione.

8. Generalizzazioni: Potenziali e Paesaggi Adattativi

Il modello di Ornstein-Uhlenbeck può essere visto come un caso speciale di una classe più generale di processi stocastici governati da potenziali. Consideriamo un'equazione della forma:

$$dX(t) = -\nabla V(X(t))\, dt + \sigma\, dW(t)$$

dove $V: \mathbb{R}^n \to \mathbb{R}$ è una funzione potenziale. Il drift è quindi il negativo del gradiente del potenziale, il che significa che il sistema tende a muoversi verso i minimi di V, analogamente a come una particella in un campo gravitazionale rotola verso i punti di minima energia potenziale.

Nel contesto linguistico, il potenziale V può essere interpretato come una funzione di "costo" o "difficoltà" articolatoria. Le regioni di basso potenziale corrispondono a configurazioni fonetiche facili da produrre e percepire, mentre le regioni di alto potenziale corrispondono a configurazioni difficili o marcate. Il cambiamento linguistico procederebbe quindi attraverso un processo di discesa del gradiente con rumore sovrapposto.

La forma del potenziale V determina la struttura del paesaggio adattativo. Potenziali con un singolo minimo danno luogo a una convergenza globale, mentre potenziali con multipli minimi separati da barriere energetiche possono portare a fenomeni di metastabilità, dove il sistema rimane intrappolato in un minimo locale per lunghi periodi prima di compiere transizioni rare ma drammatiche verso altri minimi.

Questa prospettiva fornisce un quadro unificato per comprendere sia i cambiamenti graduali sia quelli apparentemente catastrofici. I cambiamenti graduali corrispondono alla deriva all'interno di un singolo bacino attrattivo, mentre i cambiamenti catastrofici corrispondono a transizioni tra bacini diversi, innescate da fluttuazioni sufficientemente grandi da superare le barriere energetiche.

9. Tempo e Scale Temporali nel Cambiamento Linguistico

Una questione cruciale nell'applicazione dei modelli stocastici alla linguistica storica riguarda la definizione appropriata della scala temporale. Mentre nella fisica del moto browniano il tempo è misurato in secondi o millisecondi, nella linguistica storica parliamo di generazioni, secoli o millenni.

Il parametro temporale rilevante per i processi di cambiamento fonologico è probabilmente meglio concepito come il tempo generazionale piuttosto che il tempo cronologico assoluto. Ogni generazione di parlanti acquisisce il sistema fonologico attraverso l'esposizione all'input linguistico della generazione precedente, ma con possibili innovazioni e rianalisi. Il processo di trasmissione inter-generazionale introduce intrinsecamente un elemento di casualità: anche assumendo una fedeltà perfetta nella trasmissione della media della popolazione, la campionatura finita introduce varianza.

Formalmente, possiamo discretizzare il tempo in generazioni e considerare che al tempo $n+1$, la distribuzione fonetica è data da un aggiornamento stocastico della distribuzione al tempo n. Nel limite continuo, quando passiamo a considerare generazioni infinitesimamente ravvicinate, recuperiamo le equazioni differenziali stocastiche in tempo continuo.

La scala temporale caratteristica del processo è determinata dal coefficiente di diffusione σ e dal coefficiente di drift. Per il processo di Ornstein-Uhlenbeck, per esempio, il tempo di rilassamento è $\tau = 1/\theta$, che rappresenta il tempo necessario perché il sistema "dimentichi" le sue condizioni iniziali. Confrontando questo tempo caratteristico con la documentazione storica disponibile, possiamo ottenere vincoli empirici sui parametri del modello.

10. Fondamenti Probabilistici: Misure di Wiener e Spazi di Traiettorie

Dal punto di vista della teoria della misura, un processo stocastico è formalmente definito come una collezione di variabili aleatorie $\{X(t, \omega) : t \in T, \omega \in \Omega\}$, dove $(\Omega, \mathcal{F}, \mathbb{P})$ è uno spazio di probabilità, T è l'insieme dei tempi, e per ogni $\omega \in \Omega$ fissato, la funzione $t \mapsto X(t, \omega)$ è chiamata una traiettoria o realizzazione del processo.

La misura di Wiener su $\mathbb{R}^n$ è la misura di probabilità sullo spazio $C([0, \infty), \mathbb{R}^n)$ delle funzioni continue che rende il processo coordinato un processo di Wiener. L'esistenza di questa misura, dimostrata rigorosamente da Wiener, è un risultato profondo che fonda matematicamente l'intera teoria.

Per le equazioni differenziali stocastiche più generali, l'esistenza e l'unicità di soluzioni sono garantite dal teorema di esistenza e unicità, che richiede condizioni di Lipschitz sui coefficienti μ e σ. Sotto queste condizioni, esiste una misura di probabilità indotta sullo spazio delle traiettorie che è assolutamente continua rispetto alla misura di Wiener, con densità di Radon-Nikodym data dalla formula di Girsanov.

Questi aspetti tecnici, sebbene possano sembrare eccessivamente astratti, hanno implicazioni concrete per la modellizzazione linguistica. Garantiscono che i modelli proposti siano matematicamente ben posti e che le simulazioni numeriche convergano a oggetti matematici ben definiti. Inoltre, forniscono un quadro rigoroso per il confronto tra modelli diversi e per l'inferenza statistica basata su dati empirici.

La Grande Mutazione Vocalica e Fenomeni Diacronici come Processi Stocastici

1. La Grande Mutazione Vocalica Inglese: Un Caso di Studio Paradigmatico

La Grande Mutazione Vocalica (Great Vowel Shift, GVS) rappresenta uno dei fenomeni più studiati e dibattuti della linguistica storica. Questo processo, che ha interessato l'inglese tra il quindicesimo e il diciottesimo secolo, ha trasformato radicalmente il sistema vocalico della lingua, creando la caratteristica discrepanza tra ortografia e pronuncia che rende l'inglese moderno notoriamente difficile da imparare per gli stranieri.

La GVS consistette essenzialmente in una catena di spostamenti in cui le vocali lunghe alte /iː/ e /uː/ si dittongarono rispettivamente in /aɪ/ e /aʊ/, mentre le vocali medio-alte /eː/ e /oː/ si innalzarono occupando le posizioni lasciate libere, e così via lungo tutta la gerarchia vocalica. Il risultato fu una riorganizzazione sistematica dell'intero spazio vocalico, che tuttavia preservò molte delle relazioni strutturali pre-esistenti.

Dal punto di vista della modellizzazione stocastica, la GVS presenta caratteristiche particolarmente interessanti che la rendono un candidato ideale per l'applicazione delle equazioni differenziali stocastiche. Primo, il processo si è svolto su una scala temporale di diversi secoli, fornendo una risoluzione temporale sufficiente per osservare traiettorie continue. Secondo, la documentazione storica, sebbene frammentaria, è sufficientemente ricca da permettere la ricostruzione approssimativa degli stati intermedi. Terzo, il carattere sistematico del processo suggerisce l'esistenza di forze di drift ben definite, mentre la presenza di variabilità regionale e sociale indica significativi effetti di diffusione.

2. Modellizzazione del GVS attraverso SDE in Spazio Bidimensionale

Per modellizzare formalmente la GVS, consideriamo lo spazio vocalico bidimensionale parametrizzato dalle prime due formanti normalizzate. Definiamo coordinate normalizzate (x_1, x_2) dove x_1 rappresenta l'altezza (inversamente correlata con F_1) e x_2 rappresenta l'anteriorità (correlata con F_2). In queste coordinate, una vocale alta ha x_1 grande, una vocale anteriore ha x_2 grande.

Postuliamo che ciascuna vocale /V/ evolva secondo un'equazione differenziale stocastica della forma:

$$dX^V(t) = \mu^V(X^V, \{X^W\}_{W\neq V}, t)\, dt + \sigma^V\, dW^V(t)$$

dove $X^V(t) = (x_1^V(t), x_2^V(t))$ rappresenta la posizione della vocale V al tempo t, μ^V è il vettore di drift che può dipendere non solo dalla posizione della vocale stessa ma anche dalle posizioni delle altre vocali del sistema (indicato dalla dipendenza da $\{X^W\}_{W\neq V}$), e $W^V(t)$ è un processo di Wiener bidimensionale indipendente per ciascuna vocale.

La forma funzionale del drift μ^V può essere motivata da considerazioni fonologiche. Un modello ragionevole potrebbe essere:

$$\mu^V = -\nabla U(X^V) - \Sigma_{W \neq V} \nabla R(X^V, X^W)$$

dove U(X) è un potenziale individuale che spinge ogni vocale verso configurazioni articolatorie ottimali, e R(X^V, X^W) è un potenziale repulsivo che mantiene le vocali foneticamente distinte, evitando la neutralizzazione.

Il potenziale individuale U potrebbe avere la forma:

$$U(X) = k_1 (x_1 - x_1^{opt})^2 + k_2 (x_2 - x_2^{opt})^2$$

che rappresenta un'attrazione verso una configurazione ottimale (x_1^{opt}, x_2^{opt}) determinata da vincoli articolatori. Il potenziale repulsivo potrebbe essere modellizzato come:

$$R(X^V, X^W) = -c / ||X^V - X^W||^\alpha$$

per qualche costante $c > 0$ e esponente $\alpha > 0$, generando una forza repulsiva che diventa forte quando due vocali si avvicinano troppo nello spazio fonetico.

3. Dinamiche di Catena: Push Chains e Drag Chains

Una questione centrale nel dibattito sulla GVS riguarda la natura causale delle relazioni tra gli spostamenti delle diverse vocali. Due ipotesi principali sono state avanzate: la push chain hypothesis, secondo cui il dittongamento delle vocali alte ha "spinto" le vocali sottostanti ad innalzarsi per mantenere le distinzioni, e la drag chain hypothesis, secondo cui l'innalzamento delle vocali medio-alte ha "trascinato" le vocali alte verso l'alto, causando eventualmente il loro dittongamento.

Nel framework delle equazioni differenziali stocastiche accoppiate, queste due ipotesi corrispondono a diverse strutture causali nel termine di interazione. In una push chain, il drift di una vocale V dipende primariamente dalle vocali situate sopra di essa nello spazio vocalico:

$$\mu_1^V \propto \Sigma_{W:\, x_1^W > x_1^V} f(x_1^W - x_1^V)$$

dove f è una funzione monotona decrescente che descrive come la pressione esercitata da vocali più alte diminuisce con la distanza.

In una drag chain, la dipendenza è opposta:

$$\mu_1^V \propto \Sigma_{W:\, x_1^W < x_1^V} g(x_1^V - x_1^W)$$

L'analisi dei dati storici attraverso tecniche di inferenza per SDE potrebbe in linea di principio discriminare tra questi due meccanismi, esaminando quali pattern di correlazione temporale tra gli spostamenti sono più compatibili con quale struttura di interazione.

4. Parametrizzazione Temporale e Tassi di Cambiamento

I dati storici sulla GVS, ricostruiti principalmente attraverso l'analisi di rime poetiche, ortografie devianti, e descrizioni grammaticali dell'epoca, permettono di tracciare grossolanamente la cronologia degli spostamenti. Per esempio, evidenze suggeriscono che il dittongamento di /iː/ era già in corso nel quattordicesimo secolo, mentre l'innalzamento di /eː/ divenne prominente nel quindicesimo secolo.

Possiamo calibrare il modello stocastico confrontando le velocità di cambiamento previste con quelle osservate. La velocità istantanea di cambiamento è data dal drift:

$v(t) = ||\mu(X(t), t)||$

che ha dimensioni di unità fonetiche per unità di tempo. Se misuriamo lo spostamento in termini di percentuale dello spazio vocalico e il tempo in generazioni, potremmo trovare tipicamente velocità dell'ordine di $v \approx 0{,}01 - 0{,}05$ per generazione per i cambiamenti più rapidi della GVS.

Il coefficiente di diffusione σ può essere stimato dalla variabilità sincronica osservata in ciascun periodo. Studi dialettologici mostrano che in qualsiasi momento dato, la realizzazione di una vocale varia considerevolmente tra parlanti e contesti, con deviazioni standard tipicamente dell'ordine del 5-10% dello spazio vocalico totale. Questo fornisce una stima diretta di σ.

Il rapporto tra drift e diffusione determina quanto "direzionale" appare il processo. Se $||\mu|| >> \sigma$, il cambiamento appare molto sistematico e prevedibile. Se $||\mu|| \approx \sigma$, c'è sostanziale stocasticità e il cambiamento appare più erratico.

5. Soluzioni Numeriche e Simulazione Monte Carlo

Data la complessità delle equazioni differenziali stocastiche accoppiate che descrivono l'intero sistema vocalico, soluzioni analitiche esatte sono generalmente intrattabili. Dobbiamo quindi ricorrere a metodi numerici di integrazione. Il metodo standard è lo schema di Eulero-Maruyama, un'estensione stocastica del metodo di Eulero per ODE.

Discretizzando il tempo in passi Δt, lo schema è:

$X^V(t + \Delta t) = X^V(t) + \mu^V(X^V(t), \{X^W(t)\}_{W \neq V}, t)\, \Delta t + \sigma^V \sqrt{\Delta t}\, Z^V$

dove Z^V è un vettore di variabili casuali normali standard indipendenti. Questo schema converge in media quadratica alla soluzione vera quando $\Delta t \to 0$, con ordine di convergenza 1/2 in generale (ordine 1 in casi speciali).

Per simulare l'evoluzione dell'intero sistema vocalico durante la GVS, potremmo procedere come segue:

1. Inizializzare il sistema con le posizioni fonetiche del medio inglese (circa 1400)
2. Per ogni passo temporale da t = 0 a t = T (dove T corrisponde a circa 400 anni):
    - Calcolare i drift μ^V per ogni vocale V basandosi sulle posizioni correnti
    - Generare incrementi browniani indipendenti
    - Aggiornare tutte le posizioni usando lo schema di Eulero-Maruyama

3. Ripetere la simulazione molte volte per ottenere un ensemble di traiettorie

L'analisi dell'ensemble risultante fornisce predizioni statistiche: possiamo calcolare la distribuzione prevista delle vocali in qualsiasi epoca, la probabilità di particolari sequenze temporali di eventi, e la sensibilità del risultato finale alle condizioni iniziali e ai parametri.

6. Confronto con Altri Grandi Mutamenti Vocalici

La GVS inglese non è un evento isolato. Processi analoghi sono stati documentati in altre lingue germaniche e non solo. L'hochdeutsche Lautverschiebung (seconda mutazione consonantica del tedesco), pur riguardando principalmente consonanti, presenta caratteristiche strutturali simili di cambiamento a catena. In finlandese, si è verificato un innalzamento sistematico delle vocali in alcune varietà.

Confrontando questi diversi processi attraverso il framework SDE, possiamo investigare se esistono parametri universali che governano tali dinamiche. Per esempio, potremmo ipotizzare che il rapporto tra forze repulsive e forze attrattive abbia valori tipici simili attraverso diverse lingue, riflettendo vincoli cognitivi universali sulla struttura dei sistemi fonologici.

Un'analisi comparativa richiederebbe di:

1. Ricostruire gli spazi fonetici pre- e post-mutamento per ciascuna lingua
2. Stimare i parametri delle SDE (drift e diffusione) per ciascun caso
3. Testare se i parametri si raggruppano attorno a valori caratteristici
4. Investigare quali fattori (dimensione della popolazione di parlanti, contatto linguistico, presenza di ortografia standardizzata, ecc.) correlano con variazioni nei parametri

7. Il Ruolo della Variazione Sociolinguistica

Un aspetto essenziale che complica la modellizzazione SDE è la struttura sociolinguistica della variazione. Il cambiamento linguistico non procede uniformemente attraverso tutta la popolazione di parlanti, ma mostra tipicamente stratificazione per età, classe sociale, genere, e rete sociale.

Il modello teorico più influente in questo dominio è quello di Labov, che distingue tra cambiamenti dall'alto (prestigio) e cambiamenti dal basso (inconsci). Questi corrispondono a diverse strutture di diffusione nella popolazione. Un cambiamento dall'alto potrebbe essere modellizzato come un processo che inizia nelle classi sociali alte e diffonde verso il basso, mentre un cambiamento dal basso emerge nelle classi lavoratrici e può rimanere sotto il livello di consapevolezza per lungo tempo.

Per incorporare questa struttura nel framework SDE, potremmo estendere lo spazio degli stati includendo una coordinata aggiuntiva s che rappresenta la classe sociale o altra variabile stratificatrice. L'equazione diventa:

$$d[X(t), S(t)] = \mu([X, S], t)\, dt + \sigma([X, S], t)\, dW(t)$$

dove ora il drift e la diffusione dipendono anche dalla posizione sociale. Inoltre, potremmo assumere che la coordinata sociale S evolva molto più lentamente della coordinata fonetica X, riflettendo la maggiore stabilità della stratificazione sociale rispetto alla variazione linguistica.

8. Lenizione e Fortizione: Processi Direzionali con Reversioni

Mentre la GVS rappresenta un esempio di riorganizzazione massiva ma essenzialmente unidirezionale del sistema fonologico, altri processi mostrano pattern più complessi di avanzamento e retrocessione. La lenizione consonantica, cioè l'indebolimento articolatorio delle consonanti, è ampiamente attestata come tendenza diacronica in molte famiglie linguistiche.

Nelle lingue romanze, per esempio, le occlusive sorde intervocaliche latine hanno subito percorsi evolutivi diversi: in spagnolo, /p t k/ latine diventano generalmente /b d g/ (lenizione), mentre in francese il processo è andato oltre, con /t/ e /k/ che in certi contesti sono completamente scomparsi. Tuttavia, esistono anche casi di fortizione (rafforzamento), specialmente in posizioni prosodicamente prominenti.

Possiamo modellizzare questo fenomeno attraverso una SDE su uno spazio unidimensionale di "forza articolatoria" x, con:

$dx(t) = [f(x, \text{contesto}) - \gamma x] dt + \sigma dW(t)$

dove f(x, contesto) rappresenta pressioni dipendenti dal contesto (posizione nella parola, velocità di eloquio, ecc.) e $-\gamma x$ è un termine di selezione stabilizzante che penalizza deviazioni troppo estreme in entrambe le direzioni.

Il potenziale associato potrebbe avere multipli minimi locali corrispondenti a diversi "gradi di lenizione" stabili: occlusiva sorda, occlusiva sonora, fricativa, approssimante, zero (cancellazione). Le transizioni tra questi stati corrisponderebbero a salti indotti da fluttuazioni sufficientemente grandi da superare le barriere energetiche tra i minimi.

9. Assimilazioni e Coarticulazioni: Diffusione Locale nello Spazio Fonetico

Le assimilazioni rappresentano processi in cui un segmento fonetico diventa più simile a un segmento adiacente. Per esempio, /n/ diventa [m] prima di /p/ in molte lingue (assimilazione del luogo di articolazione). Questi processi sono ubiqui e rappresentano una fonte importante di variazione fonetica che può stabilizzarsi diacronicamente.

Dal punto di vista della modellizzazione SDE, le assimilazioni possono essere viste come effetti di "accoppiamento locale" tra segmenti adiacenti in una sequenza. Se $X_1(t)$ e $X_2(t)$ rappresentano le posizioni fonetiche di due segmenti consecutivi, potremmo scrivere:

$dX_1(t) = [-\nabla V(X_1) + \kappa(X_2 - X_1)] dt + \sigma dW_1(t)$ $dX_2(t) = [-\nabla V(X_2) + \kappa(X_1 - X_2)] dt + \sigma dW_2(t)$

dove $\kappa > 0$ è una costante di accoppiamento che misura la forza della tendenza assimilativa. Questo sistema tende a portare X_1 e X_2 verso configurazioni simili, pur mantenendo ciascuno ancorato al proprio attrattore individuale definito da V.

La dinamica risultante dipende dal rapporto tra la forza di accoppiamento κ e la forza di attrazione verso i valori fonologicamente specificati. Se κ è piccolo, i segmenti mantengono valori distintivi con solo piccole perturbazioni assimilative. Se κ è grande, possono verificarsi neutralizzazioni complete.

10. Metaplasmo e Cambiamenti Analogici

I cambiamenti analogici rappresentano una categoria diversa di innovazione linguistica, in cui forme irregolari vengono regolarizzate sulla base di pattern maggioritari. Un esempio classico è il livellamento analogico delle forme verbali irregolari in inglese (helped al posto del antico holp).

Questi processi sono meno naturalmente modellizzabili come diffusioni continue nello spazio fonetico, poiché coinvolgono rianalisi categoriali piuttosto che spostamenti graduali. Tuttavia, possiamo concepirli come transizioni tra bacini attrattivi discreti in uno spazio morfologico astratto.

Un modello possibile prevede un potenziale $V(x)$ con multipli minimi corrispondenti a diversi paradigmi morfologici. La profondità relativa di questi minimi riflette la "forza" o frequenza d'uso dei rispettivi paradigmi. Un paradigma minoritario corrisponde a un minimo meno profondo, e quindi più suscettibile di abbandono a favore di minimi più profondi quando perturbazioni stocastiche sono sufficientemente grandi.

La probabilità di transizione dal paradigma A al paradigma B può essere approssimativamente calcolata usando la teoria delle grandi deviazioni, che fornisce:

$$P(A \to B) \propto \exp[-\Delta V / (\sigma^2/2)]$$

dove ΔV è la barriera energetica tra i due minimi. Questo spiega perché cambiamenti analogici tendano a essere "quantici" piuttosto che graduali: sono transizioni rare tra stati discreti, innescate da fluttuazioni.

11. Palatizzazioni e Processi Categoriali

La palatizzazione, in cui consonanti velari diventano palatali o alveopalatali prima di vocali anteriori, rappresenta un altro processo estremamente comune transculturalmente. Per esempio, /k/ latino è diventato /tʃ/ in italiano prima di /e/ e /i/ (centum → cento).

Questo processo può essere visto come una transizione indotta da contesto in uno spazio articolatorio multidimensionale. La vocale anteriore successiva crea un "paesaggio energetico" modificato per la consonante precedente, abbassando la barriera verso configurazioni palatali.

Formalmente, potremmo modellizzare:

$dX_C(t) = -\nabla V(X_C, X_V)\, dt + \sigma\, dW(t)$

dove X_C è la posizione articolatoria della consonante, X_V è quella della vocale seguente, e $V(X_C, X_V)$ è un potenziale accoppiato che ha minimi diversi a seconda di X_V.

Quando X_V è anteriore, il minimo di V nella dimensione del luogo di articolazione consonantico si sposta anteriormente, favorendo realizzazioni palatali. Il processo di palatizzazione storica corrisponde alla graduale stabilizzazione di queste realizzazioni condizionate dal contesto come forme invarianti.

12. Ricapitolazione: La GVS come Prototipo di Sistema Dinamico Complesso

Ritornando alla Grande Mutazione Vocalica con cui abbiamo aperto questa sezione, possiamo ora apprezzare come essa incorpori molti degli aspetti discussi: accoppiamento tra variabili (le diverse vocali), dinamiche di catena, variabilità temporale, transizioni tra configurazioni stabili, e influenza della stratificazione sociale.

La modellizzazione attraverso sistemi di equazioni differenziali stocastiche accoppiate fornisce un framework unificato per comprendere tutti questi aspetti simultaneamente. Sebbene modelli semplificati catturino solo parte della complessità reale, forniscono intuizioni qualitative robuste e permettono di generare predizioni testabili.

Le simulazioni Monte Carlo basate su questi modelli possono inoltre aiutare a valutare ipotesi controverse. Per esempio, possiamo simulare scenari controfattuali: cosa sarebbe successo se la popolazione di parlanti fosse stata molto più grande (maggiore inerzia, minore effetto di drift genetico)? O se il contatto con altre lingue fosse stato più intenso (maggiore diffusione)?

Ramificazioni Linguistiche e Alberi Coalescenti

1. Dal Proto-Indoeuropeo alle Lingue Moderne: Processi di Speciazione Linguistica

La ricostruzione dell'albero genealogico delle lingue indoeuropee rappresenta uno dei risultati più impressionanti della linguistica comparativa. Partendo dal proto-indoeuropeo (PIE), parlato probabilmente nelle steppe pontico-caspiche circa 6000 anni fa, si sono sviluppate famiglie linguistiche che oggi comprendono la maggioranza delle lingue europee e molte lingue asiatiche, dall'hindi al farsi, dall'armeno alle lingue slave.

Il processo attraverso cui una lingua ancestrale si ramifica in lingue figlie è concettualmente analogo alla speciazione biologica: una popolazione di parlanti inizialmente omogenea si divide geograficamente o socialmente, la comunicazione tra i sottogruppi diminuisce, e

l'accumulo indipendente di innovazioni linguistiche porta gradualmente all'emergere di dialetti reciprocamente incomprensibili, e infine di lingue distinte.

Dal punto di vista della modellizzazione stocastica, la ramificazione linguistica può essere descritta attraverso processi di biforcazione stocastica. Il momento di split tra due lingue sorelle non è tipicamente un evento puntuale, ma un processo esteso durante il quale la differenziazione procede gradualmente. Tuttavia, per scopi modellistici, possiamo idealizzare questi processi estesi come eventi discreti di biforcazione che suddividono una popolazione in sottopopolazioni isolate.

2. Alberi Coalescenti e il Processo di Kingman

La teoria coalescente, sviluppata da John Kingman negli anni Ottanta nel contesto della genetica delle popolazioni, fornisce un quadro matematico rigoroso per descrivere la genealogia di campioni provenienti da una popolazione che evolve nel tempo. Sebbene originariamente concepita per analizzare dati genetici, la teoria si applica naturalmente anche alla linguistica storica.

Il coalescente di Kingman descrive la struttura genealogica di n linee di discendenza campionate dal presente, tracciate all'indietro nel tempo fino al loro antenato comune più recente (MRCA, Most Recent Common Ancestor). Il processo si basa su alcune assunzioni semplificanti: popolazione di dimensione costante, generazioni sovrapposte, e assenza di selezione.

In tempo continuo, il coalescente può essere descritto come un processo di morte puro (pure death process) che parte con n linee e procede all'indietro nel tempo. Il tempo fino alla prima coalescenza (quando due linee si uniscono) è distribuito esponenzialmente con tasso $\lambda_2 = n(n-1)/2N$, dove N è la dimensione effettiva della popolazione. Dopo la prima coalescenza rimangono n-1 linee, e il tempo fino alla successiva coalescenza è esponenziale con tasso $\lambda_3 = (n-1)(n-2)/2N$, e così via.

Il tempo totale fino all'MRCA di tutte le n linee è quindi:

$T_MRCA = T_2 + T_3 + ... + T$□

dove ciascun T_k è esponenziale con tasso k(k-1)/2N. Il valore atteso è:

E[T_MRCA] = 2N Σ_{k=2}^n 1/(k(k-1)) = 2N (1 - 1/n) ≈ 2N per n grande

Questo risultato implica che il tempo atteso fino all'antenato comune di un campione è circa 2N generazioni, indipendentemente dalla dimensione del campione (per campioni grandi), essendo dominato dal lungo tempo necessario per le ultime coalescenze.

3. Applicazione alla Linguistica: Interpretazione e Adattamenti

Nel contesto linguistico, le "linee di discendenza" sono lingue o dialetti, e la "popolazione" è la comunità linguistica ancestrale. La dimensione effettiva della popolazione N corrisponde al numero di "unità di trasmissione linguistica" nella comunità ancestrale, che potrebbe essere operazionalizzato come il numero di famiglie o di parlanti primari.

Tuttavia, l'applicazione diretta del coalescente di Kingman alla linguistica richiede alcune modifiche concettuali importanti. Primo, a differenza della genetica dove ogni individuo ha esattamente due genitori, la trasmissione linguistica coinvolge input da molteplici parlanti nella comunità. Questo corrisponde più a un modello di Wright-Fisher con mescolamento che a semplice riproduzione sessuata.

Secondo, le lingue possono andare estinte, un fenomeno che non ha un analogo diretto nella genealogia genetica (dove tutti gli individui presenti hanno necessariamente antenati). Le lingue estinte devono essere esplicitamente modellizzate come linee che terminano prima di raggiungere il presente.

Terzo, il contatto linguistico e il prestito creano connessioni orizzontali tra linee che dovrebbero essere verticalmente separate. Questo viola l'assunzione di evoluzione puramente arborescente e richiede estensioni alla teoria coalescente standard per incorporare reticolazioni.

4. Il Processo di Birth-Death e la Diversificazione Linguistica

Un'alternativa al coalescente, che procede all'indietro nel tempo, è usare processi di birth-death che procedono in avanti, descrivendo esplicitamente nascite (speciazioni) e morti (estinzioni) di lingue. Il più semplice è il processo di Yule, un processo di nascita puro dove ogni lingua esistente ha un tasso costante λ di speciare in due lingue figlie.

Partendo da una singola lingua al tempo $t = 0$, il numero di lingue al tempo t segue:

$N(t) \sim \text{Yule}(\lambda)$ con $E[N(t)] = e^{\lambda t}$

Questo processo genera crescita esponenziale illimitata, chiaramente non realistica per lunghi tempi. Una generalizzazione più realistica è il processo di birth-death con tassi λ (nascita) e μ (morte). La dinamica è governata dall'equazione:

$$\frac{dN(t)}{dt} = (\lambda - \mu) N(t)$$

con soluzione $N(t) = N_0 e^{(\lambda-\mu)t}$, che mostra crescita se $\lambda > \mu$, decrescita se $\lambda < \mu$, e equilibrio se $\lambda = \mu$.

Nel caso $\lambda = \mu$, il sistema è in stato critico. La distribuzione stazionaria del numero di lingue discendenti da un'unica lingua ancestrale dopo tempo t è geometrica, con media che rimane costante ma varianza che cresce linearmente con t. Questo comportamento critico è stato proposto come caratteristica dell'evoluzione linguistica reale.

5. Processi di Speciazione con Caratteri Continui

I modelli di birth-death descritti sopra trattano le lingue come entità discrete senza struttura interna. Per connetterli con i modelli SDE discussi nelle sezioni precedenti, dobbiamo incorporare l'evoluzione di caratteri continui (come posizioni fonetiche) lungo i rami dell'albero.

Un modello integrato potrebbe procedere come segue. Ciascuna lingua i è caratterizzata da uno stato X_i(t) ∈ ℝ^d che rappresenta il suo sistema fonologico (o più generalmente, un vettore di tratti linguistici). Tra eventi di speciazione, X_i evolve secondo:

$$dX_i(t) = \mu(X_i, t)\, dt + \sigma\, dW_i(t)$$

Quando avviene una speciazione al tempo τ, la lingua i si divide in due lingue figlie i_1 e i_2, con stati iniziali:

$$X_{i_1}(\tau) = X_i(\tau^-) + \delta_1 \quad X_{i_2}(\tau) = X_i(\tau^-) + \delta_2$$

dove δ_1 e δ_2 sono perturbazioni casuali che rappresentano la divergenza iniziale dovuta alla separazione geografica o sociale. Dopo la speciazione, le due lingue figlie evolvono indipendentemente secondo le proprie SDE.

Il tasso di speciazione λ può essere reso dipendente dallo stato, riflettendo l'idea che popolazioni in certe condizioni (grandi dimensioni, espansione territoriale rapida) hanno maggiore probabilità di frammentarsi:

$$\lambda(X, N) = \lambda_0\, f(X)\, g(N)$$

dove f(X) cattura dipendenze dal sistema linguistico e g(N) cattura dipendenze dalla dimensione della popolazione parlante.

6. Tempo di Divergenza e Tassi di Cambiamento: Metodi Glottocronologici

La glottocronologia, proposta da Morris Swadesh negli anni Cinquanta, tentava di stimare il tempo di divergenza tra lingue imparentate basandosi sulla percentuale di vocabolario condiviso, assumendo un tasso costante di rimpiazzo lessicale. Sebbene la versione originale della glottocronologia sia stata ampiamente criticata per le sue assunzioni eccessivamente semplicistiche, versioni moderne più sofisticate che incorporano modelli statistici rigorosi hanno mostrato promesse.

Dal punto di vista stocastico, il vocabolario può essere modellizzato come una collezione di variabili binarie (parola conservata vs sostituita). Se assumiamo che ciascuna parola nel vocabolario di base abbia una probabilità $\mu\, dt$ di essere rimpiazzata in un intervallo di tempo dt, e che i rimpiazzi siano indipendenti tra parole, allora la frazione di vocabolario conservato decade esponenzialmente:

$$p(t) = e^{-\mu t}$$

dove μ è il tasso di sostituzione lessicale. Osservando p tra due lingue sorelle, possiamo stimare il loro tempo di divergenza come:

$$t = -\ln(p) / \mu$$

Il problema critico è che μ non è costante: varia tra lingue, tra classi semantiche, e nel tempo. Approcci bayesiani moderni, come quelli implementati in software come BEAST,

trattano sia μ sia t come variabili aleatorie da stimare simultaneamente dai dati, usando prior informativi quando disponibili (per esempio, da evidenze archeologiche o epigrafiche).

7. La Famiglia Indoeuropea: Topologia dell'Albero e Datazione degli Split

L'albero genealogico delle lingue indoeuropee è stato oggetto di intensa ricerca per oltre due secoli. Sebbene la topologia generale sia ampiamente accettata (con ramificazioni principali in anatolico, tocharico, indo-iranico, greco, armeno, albanese, baltico, slavo, germanico, celtico, italico), questioni di dettaglio rimangono controverse, specialmente riguardo all'ordine relativo di ramificazione dei rami occidentali.

Metodi filogenetici computazionali moderni, adattati dalla biologia evoluzionistica, hanno permesso analisi quantitative. Gray e Atkinson (2003) hanno applicato metodi bayesiani a un dataset di cognati indo-europei, ottenendo stime di circa 8700 anni per il proto-indoeuropeo, significativamente più vecchio delle stime tradizionali di circa 6000 anni. Questo ha alimentato il dibattito tra l'ipotesi "kurganica" (origine nelle steppe) e l'ipotesi "anatolica" (origine in Anatolia con diffusione agricola).

Dal punto di vista della modellizzazione stocastica, possiamo trattare la topologia dell'albero come una variabile aleatoria la cui distribuzione posteriore condiziona sui dati osservati (cognati, innovazioni fonologiche condivise, ecc.). Metodi MCMC (Markov Chain Monte Carlo) permettono di campionare da questa distribuzione posteriore, fornendo non solo una stima puntuale dell'albero più probabile, ma anche misure di incertezza su ciascun nodo.

8. Estinzioni Linguistiche e Processi di Morte

L'estinzione linguistica è un fenomeno tragicamente comune nella storia umana, e drammaticamente accelerato in epoca moderna. Si stima che circa il 90% delle lingue attualmente parlate sarà estinto entro la fine di questo secolo se le tendenze attuali continuano.

Dal punto di vista modellistico, le estinzioni possono essere incorporate come un processo di morte che rimuove linee dall'albero. Il tasso di estinzione μ può dipendere da vari fattori: dimensione della popolazione parlante, prestigio socioeconomico, politiche linguistiche, grado di isolamento, ecc.

Un modello semplice potrebbe assumere che il tasso di estinzione per una lingua con N parlanti sia:

$\mu(N) = \mu_0 / (1 + N/N_0)^{\alpha}$

dove μ_0 è un tasso base, N_0 è una scala caratteristica, e $\alpha > 0$ controlla quanto rapidamente il rischio diminuisce con la dimensione della popolazione. Questo riflette l'osservazione che lingue con pochissimi parlanti ($N < 1000$) sono in pericolo critico, mentre lingue con milioni di parlanti sono relativamente sicure.

Il processo stocastico risultante è un processo di birth-death dipendente dallo stato, dove sia i tassi di nascita sia i tassi di morte dipendono dalla dimensione e da altre caratteristiche delle popolazioni linguistiche.

9. Modelli Coalescenti con Struttura di Popolazione

Il coalescente di Kingman standard assume una popolazione panmittica ben mescolata. Tuttavia, le popolazioni umane reali sono strutturate geograficamente e socialmente. Modelli coalescenti strutturati estendono la teoria base per incorporare suddivisione della popolazione e migrazione limitata.

Il modello più semplice è quello delle isole di Wright: la popolazione è suddivisa in d sottopopolazioni (isole), ciascuna di dimensione N. All'interno di ogni isola, il mescolamento è completo, ma tra isole c'è un tasso di migrazione m per generazione. Il coalescente strutturato risultante mostra due scale temporali: una scala rapida per coalescenze all'interno di isole, e una scala lenta per coalescenze che coinvolgono linee in isole diverse.

Il tempo atteso fino alla coalescenza di due linee campionate dalla stessa isola è circa 2N/d, mentre il tempo atteso per linee da isole diverse dipende criticamente dal tasso di migrazione. Nel limite di migrazione nulla, le linee possono coalescere solo dopo che per deriva genetica una si è spostata nell'isola dell'altra, che richiede tempo dell'ordine di 1/m.

Nel contesto linguistico, questo modello è rilevante per situazioni in cui una famiglia linguistica è frammentata in dialetti geograficamente separati con scambi limitati. I dialetti dell'italiano prima dell'unificazione, per esempio, mostravano esattamente questa struttura.

10. Reticolazioni e Contatto Linguistico: Oltre gli Alberi

Una limitazione fondamentale dei modelli puramente arborei è che ignorano il contatto linguistico e il prestito. Nella realtà, lingue contigue influenzano reciprocamente i loro sistemi fonologici, morfologici e lessicali, creando somiglianze che non riflettono discendenza comune ma convergenza.

Le reti filogenetiche (phylogenetic networks) estendono gli alberi permettendo bordi orizzontali che rappresentano eventi di ibridazione o prestito. Formalmente, una rete è un grafo diretto aciclico con una singola radice, dove alcuni nodi hanno due genitori (corrispondendo a eventi di ibridazione) piuttosto che uno solo.

L'inferenza di reti da dati linguistici è computazionalmente più difficile che l'inferenza di alberi, poiché lo spazio delle possibili reti è molto più grande. Tuttavia, metodi come NeighborNet e SplitsTree hanno mostrato successo nell'identificare segnali di reticolazione in dataset linguistici.

Dal punto di vista della modellizzazione stocastica, possiamo incorporare eventi di prestito come jump processes che trasferiscono istantaneamente caratteri linguistici da una linea a un'altra. Se $\lambda_{transfer}$ è il tasso di eventi di prestito per unità di tempo, e assumiamo che ciascun evento trasferisca una frazione θ dei caratteri, allora l'evoluzione combinata è descritta da una SDE con drift, diffusione, e salti:

$$dX_i(t) = \mu\, dt + \sigma\, dW_i(t) + \Sigma_j\, \theta_{ij}\, (X_j - X_i)\, dN_{ij}(t)$$

dove $N_{ij}(t)$ sono processi di Poisson indipendenti con tasso λ_{ij} che governano gli eventi di prestito dalla lingua j alla lingua i.

11. Diversità Linguistica e Modelli Neutrali vs Selettivi

Un dibattito centrale nella linguistica evolutiva riguarda il ruolo relativo della deriva neutrale rispetto alla selezione adattativa. In genetica delle popolazioni, la teoria neutrale di Kimura postula che la maggior parte della variazione genetica sia selettivamente neutra, evolvendosi primariamente per deriva genetica. Esiste un analogo linguistico?

La deriva neutrale nel contesto linguistico significherebbe che la maggior parte dei cambiamenti linguistici non conferisce né vantaggio né svantaggio comunicativo, ma si fissa semplicemente per campionamento stocastico attraverso le generazioni. Al contrario, una prospettiva selettiva enfatizzerebbe che i cambiamenti linguistici sono guidati da pressioni funzionali: economia articolatoria, distintività percettiva, facilità di acquisizione, ecc.

La realtà è probabilmente una combinazione. Alcuni cambiamenti sembrano chiaramente neutri (come la scelta arbitraria tra diverse forme sinonime), mentre altri sembrano adattativi (come semplificazioni morfologiche in lingue con molti apprendenti adulti). La modellizzazione SDE può incorporare entrambi gli aspetti: il drift μ rappresenta pressioni selettive, mentre la diffusione σ rappresenta deriva neutrale.

Il rapporto $Ne = ||\mu||^2 / \sigma^2$ può essere interpretato come una "dimensione effettiva di popolazione" che misura quanto la dinamica è dominata da selezione versus deriva. Per $Ne >> 1$, la selezione domina e il sistema converge deterministicamente verso configurazioni ottimali. Per $Ne << 1$, la deriva domina e il sistema vaga casualmente nello spazio degli stati.

12. Hotspots di Diversità e Refugia Linguistici

Un pattern notevole nella diversità linguistica mondiale è la sua estrema non-uniformità geografica. Alcune regioni, come la Nuova Guinea, ospitano centinaia di lingue in aree relativamente piccole, mentre vaste regioni di Eurasia sono linguisticamente più omogenee. Questo pattern richiede spiegazione.

Una prospettiva modellistica suggerisce che la diversità è massimizzata quando il tasso di speciazione eccede il tasso di estinzione, e quando la connettività tra sottopopolazioni è bassa. Regioni montagnose e isolate insulari favoriscono entrambe queste condizioni: la frammentazione topografica facilita isolamento e divergenza, mentre la bassa mobilità riduce il livellamento per contatto.

Possiamo modellizzare questo attraverso SDE in spazi con ostacoli e barriere. Supponiamo che lo spazio geografico G contenga regioni di alta e bassa permeabilità linguistica. La diffusione effettiva è allora eterogenea:

$$\sigma(x) = \sigma_0 / (1 + h(x))$$

dove $h(x)$ è una funzione di "difficoltà" che è grande in montagne, mari, deserti, ecc. Questo porta a clustering e speciazione aumentata nelle regioni ad alta $h(x)$.

Le simulazioni di tali modelli spazialmente espliciti riproducono qualitativamente i pattern osservati di diversità linguistica, suggerendo che considerazioni geografiche e demografiche

semplici possono spiegare gran parte della variazione globale senza invocare differenze culturali intrinseche.

Prestiti Lessicali come Jump Processes

1. Discontinuità nell'Evoluzione Linguistica: Il Ruolo dei Salti

Mentre i processi di diffusione continua descritti dalle equazioni differenziali stocastiche catturano bene l'evoluzione graduale dei sistemi fonologici e morfologici, molti fenomeni linguistici sono intrinsecamente discontinui. I prestiti lessicali rappresentano il caso paradigmatico: una parola straniera viene adottata all'improvviso da una comunità linguistica, introducendo un salto discreto nel lessico che non può essere derivato da trasformazioni graduali del materiale lessicale esistente.

Matematicamente, questi fenomeni richiedono l'introduzione di processi con salti (jump processes) che complementano la componente di diffusione continua. Un processo con salti è caratterizzato da traiettorie che sono costanti a tratti, con discontinuità improvvise a tempi aleatori. Il framework generale per tali processi è fornito dai processi di Lévy, che generalizzano il moto browniano permettendo salti oltre alla diffusione continua.

La teoria dei processi di Lévy si basa sul teorema di rappresentazione di Lévy-Khintchine, che caratterizza completamente la distribuzione di un processo di Lévy attraverso tre componenti: un drift deterministico, una componente di diffusione browniana, e una misura di Lévy che specifica la distribuzione dei salti. Per un processo di Lévy L(t) in $\mathbb{R}^d$, la funzione caratteristica ha la forma:

$$E[e^{i\langle\xi,L(t)\rangle}] = \exp[t(i\langle b,\xi\rangle - (1/2)\langle\xi,A\xi\rangle + \int_{\mathbb{R}^d} (e^{i\langle\xi,x\rangle} - 1 - i\langle\xi,x\rangle 1_{\{||x||<1\}}) \nu(dx))]$$

dove b è il drift, A è la matrice di covarianza della componente browniana, e ν è la misura di Lévy che governa i salti. La misura ν deve soddisfare $\int \min(1, ||x||^2)\, \nu(dx) < \infty$ per garantire che il processo sia ben definito.

2. Processi di Poisson Composti: Modello Base per Prestiti

Il processo di Lévy più semplice con salti è il processo di Poisson composto, definito come:

$$X(t) = X(0) + \Sigma_{i=1}^{N(t)} Y_i$$

dove N(t) è un processo di Poisson con tasso λ, e $\{Y_i\}$ è una sequenza i.i.d. di variabili casuali che rappresentano le dimensioni dei salti, indipendente da N(t).

Nel contesto dei prestiti lessicali, N(t) rappresenta il numero di eventi di prestito che sono avvenuti fino al tempo t, mentre Y_i rappresenta l'"effetto linguistico" dell'i-esimo prestito. Quest'ultimo potrebbe essere un vettore multidimensionale che cattura vari aspetti: la

distanza fonetica della parola prestata dal sistema ricevente, il numero di fonemi introdotti, l'impatto sulla struttura sillabica, l'area semantica, ecc.

Il tasso λ del processo di Poisson dipende criticamente dall'intensità del contatto linguistico. Per lingue in contatto intenso (bilingualismo diffuso, immigrazione massiccia, dominazione politica), λ può essere molto grande, mentre per lingue isolate λ è prossimo a zero. Possiamo modellizzare:

$\lambda = \lambda_0 \cdot f(\text{contatto}) \cdot g(\text{prestigio_relativo})$

dove f quantifica l'intensità del contatto e g cattura il prestigio relativo delle lingue, poiché i prestiti tendono a fluire da lingue di maggiore prestigio a lingue di minore prestigio.

3. Distribuzione delle Dimensioni dei Salti: Leggi di Potenza

Un aspetto empiricamente interessante riguarda la distribuzione delle "dimensioni" dei prestiti lessicali. Se misuriamo la dimensione come la frequenza d'uso della parola prestata, o come il numero di derivati che genera, osserviamo distribuzioni altamente asimmetriche con code pesanti: molti prestiti hanno impatto minimo, mentre pochi hanno impatto enorme.

Questo suggerisce l'uso di distribuzioni a coda pesante per i salti Y_i. Una scelta naturale è una distribuzione di Pareto:

$P(Y > y) = (y_0/y)^{\alpha}$ per $y \geq y_0$

dove $\alpha > 0$ è l'esponente di Pareto. Per $\alpha \in (1,2)$, la distribuzione ha varianza infinita, portando a un processo di Lévy chiamato processo α-stabile. Questi processi mostrano auto-similarità e sono caratterizzati da fluttuazioni estreme occasionali.

L'uso di processi α-stabili per modellizzare prestiti cattura l'idea che l'impatto del contatto linguistico può variare enormemente: la maggior parte degli eventi di contatto produce piccole perturbazioni, ma occasionalmente si verificano superstratum massivi (come l'influenza normanna sull'inglese o l'influenza araba sullo spagnolo) che trasformano radicalmente la lingua ricevente.

4. Equazioni Integro-Differenziali Stocastiche

Quando combiniamo diffusione continua e salti, otteniamo equazioni integro-differenziali stocastiche (jump-diffusion SDEs) della forma generale:

$dX(t) = \mu(X(t^-), t)\, dt + \sigma(X(t^-), t)\, dW(t) + \int_{\mathbb{R}^d} z\, \tilde{N}(dt, dz)$

dove $\tilde{N}(dt, dz)$ è una misura di Poisson compensata associata alla misura di Lévy ν. Il termine integrale cattura l'effetto cumulativo di tutti i salti di dimensione z che avvengono nell'intervallo dt.

L'analogo dell'equazione di Fokker-Planck per processi con salti è l'equazione integro-differenziale di Kolmogorov forward:

$\partial p/\partial t = -\nabla \cdot [\mu p] + (1/2)\nabla \cdot \nabla \cdot [Dp] + \int_{\mathbb{R}^d} [p(x-z,t)\nu(z) - p(x,t)\nu(z)]\, dz$

Il termine integrale aggiuntivo descrive come la probabilità in x cambia a causa di salti: il primo termine rappresenta arrivi in x da x-z, il secondo rappresenta partenze da x verso x+z.

Questa equazione governa l'evoluzione della distribuzione di probabilità di variabili linguistiche soggette sia a cambiamento graduale (drift e diffusione) sia a innovazioni improvvise (salti).

5. Stratificazione Lessicale: Vocabolario di Base vs Vocabolario Culturale

Una distinzione fondamentale nella lessicologia è quella tra vocabolario di base (core vocabulary) e vocabolario culturale. Il vocabolario di base comprende termini per concetti universali (parti del corpo, numeri bassi, pronomi, azioni fondamentali) ed è notevolmente resistente al prestito. Il vocabolario culturale (tecnologia, cucina, istituzioni, ecc.) è invece altamente suscettibile a prestiti.

Possiamo modellizzare questa distinzione attraverso tassi di salto eterogenei. Suddividiamo il lessico in strati L_1, L_2, ..., L□ ordinati per "resilienza", e assumiamo che il tasso di prestito per lo strato i sia:

$\lambda_i = \lambda_0 \cdot r^i$

dove $0 < r < 1$ è un fattore di riduzione. Questo genera una gerarchia esponenziale dove gli strati più profondi (vocabolario di base) hanno tassi di prestito esponenzialmente più bassi.

I dati empirici dalla lista di Swadesh (una lista standard di 100-200 termini di vocabolario di base usata in glottocronologia) supportano questa strutturazione: i tassi di sostituzione osservati per termini di base sono tipicamente dell'ordine di 10-20% per millennio, molto più bassi dei tassi per vocabolario specializzato che possono raggiungere 50-80% per secolo in situazioni di contatto intenso.

6. Adattamento Fonologico dei Prestiti: Un Processo Stocastico a Due Stadi

Quando una parola viene prestata, non entra generalmente nella lingua ricevente nella sua forma fonologica originale, ma subisce adattamenti per conformarsi alla fonologia nativa. Questo processo può essere modellizzato come stocastico, poiché diverse strategie di adattamento sono possibili e la scelta dipende da fattori variabili.

Consideriamo un modello a due stadi:

1. **Stadio di Percezione**: Il parlante ricevente percepisce la forma straniera X_straniera attraverso un filtro percettivo che la mappa nello spazio fonetico nativo con rumore:

   $X_{percepita} = M(X_{straniera}) + \sigma_{percezione} \cdot \varepsilon$

   dove M è una mappa di categoria (che proietta fonemi stranieri sulle categorie native più vicine) e ε è rumore gaussiano.

2. **Stadio di Produzione**: Il parlante produce una forma conformata alla fonologia nativa risolvendo un problema di ottimizzazione stocastica:

   $$X_{prodotta} = \text{argmin}_X [d(X, X_{percepita}) + \lambda \cdot C(X)] + \sigma_{produzione} \cdot \delta$$

   dove d è una distanza fonetica, C è una funzione di costo che penalizza violazioni dei vincoli fonologici nativi, λ controlla il trade-off tra fedeltà e ben-formatezza, e δ è rumore nella produzione.

Il risultato complessivo è un mapping stocastico $X_{straniera} \rightarrow X_{adattata}$ che dipende dai parametri $\sigma_{percezione}$, $\sigma_{produzione}$, e λ. La variabilità negli adattamenti osservati empiricamente riflette la variabilità in questi parametri attraverso parlanti e contesti.

7. Cascade Effects: Prestiti che Innescano Ristrutturazioni

Un fenomeno particolarmente interessante è quello dei prestiti che innescano cambiamenti a catena nel sistema ricevente. L'introduzione massiccia di prestiti con particolari caratteristiche fonologiche può portare alla fonologizzazione di contrasti precedentemente allofoni, o all'introduzione di nuovi fonemi.

L'inglese fornisce esempi classici: i prestiti francesi dopo la conquista normanna hanno introdotto fricative sonore /v z ʒ/ che erano inizialmente solo allofoni delle sorde /f s ʃ/. L'accumulo di coppie minimali (vine vs fine, rouge vs rush) ha fonologizzato questi contrasti.

Possiamo modellizzare questo come un processo a soglia: il sistema fonologico ha una certa resilienza, ma quando il numero di prestiti contenenti un contrasto non-nativo supera una soglia critica θ_c, avviene una transizione di fase verso un nuovo sistema fonologico che incorpora quel contrasto nativamente.

Formalmente, se $N_{contrasto}(t)$ è il numero cumulativo di prestiti con un certo contrasto, e θ_c è la soglia critica, allora:

$$P(\text{fonologizzazione} \mid N_{contrasto}) = 1 / (1 + e^{-k(N_{contrasto} - \theta_c)})$$

dove k controlla la ripidità della transizione. Per $k \rightarrow \infty$, otteniamo una transizione sharp; per k piccolo, la transizione è graduale.

8. Relazioni Spaziali e Diffusione Areale

I prestiti non sono distribuiti uniformemente nello spazio geografico, ma mostrano forte struttura spaziale. Lingue geograficamente vicine tendono a condividere prestiti, creando aree linguistiche (Sprachbünde) caratterizzate da innovazioni condivise non dovute a parentela comune.

Possiamo estendere il modello di jump process per incorporare struttura spaziale. Supponiamo che le lingue siano distribuite nello spazio geografico, e che il tasso di prestito dalla lingua i alla lingua j sia:

$\lambda_{i\square} = \lambda_0 \cdot \exp(-d(i,j)/d_0) \cdot f(\text{prestigio_i}, \text{prestigio_j})$

dove d(i,j) è la distanza geografica e d_0 è una scala caratteristica di decadimento. Questo modello cattura l'idea che il contatto linguistico decade esponenzialmente con la distanza.

Simulando questo processo su una griglia di lingue, osserviamo l'emergere di cluster di somiglianze lessicali che riflettono sia parentela genealogica sia contatto geografico. Metodi statistici per distinguere questi due segnali sono un'area attiva di ricerca in linguistica computazionale.

9. Bilinguismo e Code-Switching: Processi di Switching Stocastico

In comunità bilingui, i parlanti alternano frequentemente tra le due lingue (code-switching), e questo processo può essere modellizzato come un processo di Markov a tempo continuo con due stati (lingua A e lingua B).

Il sistema evolve secondo una matrice generatrice Q:

$Q = [-\lambda_AB \ \lambda_AB \ ] \ [\lambda_BA \ -\lambda_BA \ ]$

dove λ_AB è il tasso di switch da A a B, e viceversa. La distribuzione stazionaria è:

$\pi_A = \lambda_BA / (\lambda_AB + \lambda_BA) \ \pi_B = \lambda_AB / (\lambda_AB + \lambda_BA)$

che rappresenta la frazione di tempo a lungo termine spesa in ciascuna lingua.

I tassi di switching dipendono da numerosi fattori sociolinguistici: argomento della conversazione, interlocutore, setting formale vs informale, competenza relativa nelle due lingue, atteggiamenti linguistici. Modellizzare questi fattori come modulatori dei tassi di switching fornisce un framework quantitativo per predire pattern di uso linguistico in comunità multilingui.

Il code-switching intenso può facilitare il prestito lessicale, poiché aumenta l'esposizione a materiale linguistico straniero. Possiamo accoppiare il modello di switching con il modello di jump process per prestiti:

$\lambda_\text{prestiti} = \lambda_\text{base} + \beta \cdot (\text{frazione_tempo_lingua_B})$

dove β quantifica quanto l'esposizione aumentata facilita l'adozione di prestiti.

10. Relexificazione e Pidginizzazione: Salti Massivi

In situazioni estreme di contatto, come nella formazione di lingue pidgin e creole, possiamo osservare fenomeni di relexificazione quasi totale, dove l'intero lessico di una lingua viene sostituito con materiale da un'altra lingua, pur mantenendo (almeno inizialmente) gran parte della struttura grammaticale originale.

Questi eventi rappresentano salti giganteschi nello spazio delle lingue possibili, e richiedono modellizzazione speciale. Una prospettiva è vederli come transizioni di fase innescate

quando l'intensità del contatto supera una soglia critica. Al di sotto della soglia, i prestiti sono incorporati gradualmente; al di sopra, il sistema collassa e si riorganizza radicalmente.

Formalmente, potremmo modellizzare un potenziale V(X, I) dove X è lo stato linguistico e I è l'intensità del contatto. Per I < I_c, il potenziale ha un singolo minimo corrispondente alla lingua nativa. Per I > I_c, emerge un nuovo minimo corrispondente a una configurazione pidginizzata/creolizzata, e il sistema può transitare verso questo nuovo minimo.

La teoria delle catastrofi di Thom fornisce un framework matematico per classificare tali transizioni. La creolizzazione corrisppponderebbe a una catastrofe fold o cusp, dove piccoli cambiamenti nel parametro di controllo (intensità di contatto) inducono cambiamenti qualitativi discontinui nel sistema.

11. Prestiti Semantici: Calcoli di Significato

Oltre ai prestiti lessicali (dove la forma fonetica è copiata), esistono prestiti semantici (calchi) dove solo il significato viene trasferito, usando materiale morfologico nativo. Per esempio, grattacielo in italiano è un calco di skyscraper inglese.

I calcoli richiedono un livello di analisi compositiva: il parlante deve scomporre il termine straniero nei suoi componenti semantici, trovare equivalenti nativi per ciascun componente, e ricomporli secondo la morfologia nativa. Questo è un processo più complesso del semplice copiare una forma fonologica.

Possiamo modellizzare il tasso di calchi come funzione della trasparenza compositiva del termine straniero:

$\lambda_calco = \lambda_0 \cdot T(termine)$

dove $T(termine) \in [0,1]$ misura quanto il termine è compositivamente trasparente. Termini opachi (molti idiomi) hanno $T \approx 0$ e raramente generano calchi, mentre termini trasparenti hanno $T \approx 1$.

Il processo di calco introduce meno perturbazione fonologica del prestito diretto, ma può comunque ristrutturare campi semantici. L'accumulo di calchi può spostare le estensioni semantiche di morfemi nativi, un tipo di cambiamento semantico indotto da contatto.

12. Dinamiche Temporali dei Prestiti: Ondate e Stratificazione

I prestiti non entrano uniformemente nel tempo, ma spesso in ondate correlate con eventi storici specifici. L'inglese mostra stratificazione chiara: uno strato di prestiti scandinavi (800-1100 d.C.), uno strato normanno-francese massiccio (1100-1300 d.C.), prestiti latini del Rinascimento (1500-1700 d.C.), e prestiti moderni da numerose lingue.

Possiamo modellizzare questo attraverso un processo di Poisson non-omogeneo dove il tasso $\lambda(t)$ varia nel tempo:

$\lambda(t) = \lambda_base + \Sigma_i A_i \exp[-(t-t_i)^2/(2\sigma_i^2)]$

dove ciascun termine gaussiano rappresenta un'ondata di prestiti centrata al tempo t_i con ampiezza A_i e durata caratteristica σ_i, corrispondente a un evento di contatto storico.

La stratificazione risultante è preservata nella lingua moderna attraverso differenze fonologiche: i prestiti più antichi sono generalmente più integrati e mostrano adattamenti più completi alla fonologia nativa, mentre prestiti recenti mantengono caratteristiche fonologiche straniere.

Questo può essere incorporato nel modello facendo sì che il grado di adattamento fonologico decresca con il tempo dalla prima adozione:

$$\text{adattamento}(t) = A_0 [1 - \exp(-t/\tau)]$$

dove τ è una costante di tempo caratteristica per l'integrazione fonologica completa, tipicamente dell'ordine di diverse generazioni.

Analogie con la Genetica delle Popolazioni - Modelli Neutrali vs Selettivi

1. La Sintesi Evolutiva: Biologia e Linguistica come Scienze Storiche

L'analogia tra evoluzione biologica ed evoluzione linguistica ha una lunga storia, risalente almeno a Charles Darwin, che nel 1871 scrisse: "La formazione di lingue differenti e di specie distinte, e le prove che entrambe si sono sviluppate attraverso un processo graduale, sono curiosamente parallele." Questa intuizione fondamentale ha trovato espressione matematica rigorosa con lo sviluppo della genetica delle popolazioni nel ventesimo secolo, i cui strumenti formali si rivelano straordinariamente applicabili anche alla linguistica storica.

Entrambe le discipline studiano sistemi che evolvono nel tempo attraverso processi di trasmissione con variazione, soggetti a forze sia deterministiche (selezione) sia stocastiche (deriva). Entrambe devono confrontarsi con il problema dell'inferenza di processi storici da dati sincronici frammentari. Entrambe operano su multiple scale temporali, dalle fluttuazioni generazionali ai pattern macroevolutivi che si dispiegano su milioni di anni (o millenni, nel caso linguistico).

La differenza cruciale è che mentre l'eredità genetica è particolata (geni discreti trasmessi con meccanismi mendeliani ben definiti), l'eredità linguistica è continua e sfumata, con bambini che acquisiscono il linguaggio da input distribuito su molti parlanti. Tuttavia, questa differenza è meno fondamentale di quanto possa sembrare: entrambi i sistemi possono essere descritti attraverso processi stocastici che operano su spazi di stati ad alta dimensionalità.

2. Il Modello di Wright-Fisher: Fondamento della Genetica delle Popolazioni

Il modello di Wright-Fisher, proposto indipendentemente da Sewall Wright e Ronald Fisher negli anni Trenta, rappresenta la pietra angolare della teoria genetica delle popolazioni. Considera una popolazione diploide di dimensione costante N, con generazioni discrete e non sovrapposte. A ciascuna generazione, i 2N gameti della generazione successiva sono scelti con rimpiazzo dalla generazione precedente.

Per un singolo locus biallelic con alleli A e a, se la frequenza di A nella generazione t è p_t, allora la frequenza nella generazione successiva è una variabile aleatoria binomiale:

$2N \cdot p_{t+1} \sim \text{Binomiale}(2N, p_t)$

Nel limite di grandi popolazioni, questo processo discreto converge a una diffusione continua governata dall'equazione differenziale stocastica:

$dp(t) = \sigma\sqrt{[p(t)(1-p(t))/(2N)]}\, dW(t)$

dove σ è la deviazione standard e W(t) è un processo di Wiener standard. Questa è l'equazione di Wright-Fisher neutra, che descrive la deriva genetica in assenza di selezione, mutazione o migrazione.

3. Trasposizione al Contesto Linguistico: Drift delle Varianti

Il parallelismo con la linguistica è immediato se pensiamo alle "varianti linguistiche" (diverse pronunce di una parola, diverse forme morfologiche, diverse costruzioni sintattiche) come analoghe agli alleli genetici. In ogni generazione, i bambini campionano il loro input linguistico dalla generazione precedente, ma con variabilità stocastica.

Consideriamo due varianti fonetiche di uno stesso fonema, V_1 e V_2 (per esempio, diverse realizzazioni di /r/ come alveolare vs uvulare). Se p_t è la frequenza della variante V_1 nella generazione t, e assumiamo campionamento neutrale, otteniamo esattamente la stessa dinamica di Wright-Fisher:

$dp(t) = \sqrt{[p(t)(1-p(t))/(2N_e)]}\, dW(t)$

dove N_e è la "dimensione effettiva della popolazione linguistica", che non corrisponde necessariamente alla dimensione demografica effettiva ma riflette il numero di "unità di trasmissione" indipendenti.

Le proprietà chiave di questo processo sono:

1. **Deriva Genetica**: In assenza di altre forze, le frequenze alleli (o varianti) vagano casualmente, con varianza che aumenta nel tempo.

2. **Fissazione Eventuale**: Con probabilità 1, alla fine una delle due varianti si fissa (raggiunge frequenza 1) e l'altra si perde, a meno che non intervengano mutazioni

che reintroducano variazione.

3. **Tempo di Fissazione**: Il tempo atteso fino alla fissazione è circa 4N_e generazioni, molto lungo per popolazioni grandi.

4. **Probabilità di Fissazione**: Per una variante neutrale con frequenza iniziale p_0, la probabilità di fissazione è semplicemente p_0 (indipendentemente da N).

4. Selezione Naturale vs Selezione Linguistica

Quando introduciamo selezione, il modello di Wright-Fisher si arricchisce. Se l'allele A ha un vantaggio selettivo s rispetto ad a, l'equazione di diffusione diventa:

$$dp(t) = s\,p(t)(1-p(t))\,dt + \sqrt{[p(t)(1-p(t))/(2N_e)]}\,dW(t)$$

Il primo termine rappresenta il drift deterministico dovuto alla selezione, il secondo la deriva stocastica. Il rapporto N_e s determina quale forza domina: per $N_e s \gg 1$, la selezione è efficace; per $N_e s \ll 1$, la deriva domina.

Nel contesto linguistico, la "selezione" corrisponderebbe a vantaggi funzionali di certe varianti. Questi potrebbero includere:

- **Facilità articolatoria**: Varianti che richiedono meno sforzo muscolare hanno un vantaggio.
- **Distintività percettiva**: Varianti che sono più facilmente distinguibili nel rumore hanno un vantaggio.
- **Prestigio sociale**: Varianti associate a gruppi di alto status hanno un vantaggio in contesti formali.
- **Frequenza**: Varianti più frequenti sono più facilmente acquisite dai bambini (frequency bias).

La sfida è quantificare questi vantaggi e stimare i coefficienti di selezione s. Studi sperimentali di percezione e produzione possono fornire stime, ma la traduzione a coefficienti evolutivi richiede modellizzazione attenta della trasmissione inter-generazionale.

5. La Teoria Neutrale di Kimura e il Dibattito Neutralismo-Selezionismo

Motoo Kimura propose nel 1968 la teoria neutrale dell'evoluzione molecolare, sostenendo che la maggior parte delle sostituzioni genetiche osservate a livello molecolare sono selettivamente neutre, fissandosi per deriva piuttosto che per selezione. Questa proposta scatenò un dibattito intenso che continua ancora oggi.

La teoria neutrale fa predizioni specifiche e testabili:

1. Il tasso di sostituzione dovrebbe essere costante nel tempo (orologio molecolare).
2. Il polimorfismo entro specie dovrebbe correlate con la divergenza tra specie.
3. La maggior parte delle sostituzioni dovrebbe avvenire in siti non funzionalmente vincolati.

Esiste un dibattito analogo in linguistica? Labov e altri sociolinguisti hanno enfatizzato che molti cambiamenti sonori non sembrano avere chiare motivazioni funzionali, suggerendo che la deriva neutrale potrebbe dominare. Altri linguisti, particolarmente nell'approccio funzionalista, sostengono che considerazioni di economia e marcatezza guidano sistematicamente il cambiamento.

Possiamo testare queste ipotesi applicando i test statistici sviluppati in genetica. Per esempio, il test di neutralità di Tajima confronta due stimatori diversi della diversità genetica che dovrebbero coincidere sotto neutralità. Deviazioni indicano selezione (o violazioni di altre assunzioni).

6. Polimorfismo Bilanciato e Variazione Sociolinguistica Stabile

Un fenomeno interessante in genetica è il polimorfismo bilanciato, dove selezione mantiene stabilmente due o più alleli nella popolazione. Il meccanismo classico è l'eterozigote advantage, dove individui eterozigoti (Aa) hanno fitness maggiore di entrambi gli omozigoti (AA e aa).

Il caso paradigmatico è l'anemia falciforme: l'allele S causa anemia in omozigoti (SS), ma eterozigoti (AS) hanno resistenza aumentata alla malaria. In regioni malariche, si stabilisce un equilibrio con entrambi gli alleli presenti.

L'analogo linguistico sarebbe variazione sociolinguistica stabile, dove due varianti coesistono a lungo termine ciascuna favorita in contesti diversi. Un esempio è la variazione tra forme standard e dialettali in comunità diglossiche: la forma standard ha vantaggio in contesti formali, la forma dialettale in contesti informali e di solidarietà ingroup.

La dinamica può essere modellizzata con selezione dipendente dalla frequenza:

$s(p) = s_0 (0,5 - p)$

dove il vantaggio selettivo cambia segno attorno a $p = 0,5$. Questo crea un attrattore stabile a $p = 0,5$, mantenendo entrambe le varianti.

7. Mutazione e Innovazione: Fonti di Variazione

In genetica, la mutazione introduce continuamente nuova variazione. Il tasso di mutazione μ è tipicamente molto basso ($\sim 10^{-8}$ - 10^{-9} per sito per generazione nei mammiferi), ma il grande numero di siti genomici significa che la mutazione complessiva è significativa.

L'equilibrio tra mutazione e deriva determina la diversità genetica stazionaria. Per il modello di infinite alleles (dove ogni mutazione crea un allele nuovo), l'eterozigosità attesa all'equilibrio è:

$H = 4N_e \mu / (1 + 4N_e \mu)$

che dipende dal prodotto $N_e \mu$, chiamato theta (θ). Popolazioni grandi o con tasso mutazionale alto hanno maggiore diversità.

In linguistica, l'"innovazione" gioca un ruolo analogo alla mutazione, introducendo nuove varianti fonetiche, morfologiche o lessicali. Tuttavia, i tassi di innovazione linguistica sono probabilmente molto più alti dei tassi mutazionali genetici, e le innovazioni non sono uniformemente distribuite (certi individui innovano più di altri).

Possiamo stimare il tasso di innovazione θ_lang da dati sincronici sulla diversità, poi usarlo per calibrare modelli evolutivi. Watterson ha proposto uno stimatore basato sul numero di varianti segreganti:

θ̂_W = S / Σ_{i=1}^{n-1} (1/i)

dove S è il numero di siti polimorfici in un campione di dimensione n. Applicato a dati dialettali, questo potrebbe fornire stime di θ_lang.

8. Migrazione e Flusso Genico vs Contatto Linguistico

La migrazione in genetica delle popolazioni introduce flusso genico tra sottopopolazioni, omogeneizzando la diversità. Il modello delle isole con migrazione simmetrica al tasso m tra d isole porta a una F_st (misura di differenziazione tra popolazioni) all'equilibrio di:

F_st ≈ 1 / (1 + 4N_e m)

Maggiore è il flusso migratorio m, minore è la differenziazione. Per m → 0, F_st → 1 (popolazioni completamente differenziate); per m → ∞, F_st → 0 (popolazione panmittica).

Il contatto linguistico gioca un ruolo analogo, omogeneizzando differenze dialettali. Possiamo definire una misura analoga F_lt di differenziazione linguistica tra dialetti:

F_lt = (diversità_totale - diversità_media_entro_dialetti) / diversità_totale

e relazionarla all'intensità del contatto. Studi empirici mostrano che F_lt per dialetti europei decresce con la facilità di comunicazione (misurata da fattori come distanza, connessione stradale, differenze economiche).

9. Coevoluzione e Epistasi: Interazioni tra Tratti

Un'area complessa in entrambe le discipline riguarda le interazioni tra caratteri. In genetica, l'epistasi descrive come l'effetto di un gene dipende da altri geni. In linguistica, molti cambiamenti sono condizionati da altri aspetti del sistema.

Per esempio, la lenizione consonantica può essere condizionata dalla struttura prosodica, che a sua volta dipende dal sistema accentuale. Cambiamenti in un componente del sistema possono catalizzare o bloccare cambiamenti in altri componenti.

Possiamo modellizzare questo attraverso fitness landscapes multidimensionali dove il fitness di una configurazione $(x_1, x_2, ..., x_n)$ non fattorizza in contributi indipendenti:

$$W(x_1, x_2, ..., x_n) \neq w_1(x_1) \cdot w_2(x_2) \cdot ... \cdot w_n(x_n)$$

ma include termini di interazione. Paesaggi con forte epistasi mostrano topologie complesse con multipli picchi separati da valli, portando a dipendenza sensibile dalle condizioni iniziali e possibile multistabilità.

10. Selezione Direzionale, Stabilizzante e Disruptiva

La teoria evoluzionistica distingue tre tipi di selezione naturale:

- **Direzionale**: favorisce valori estremi di un tratto in una direzione (esempio: aumento graduale della statura).
- **Stabilizzante**: favorisce valori intermedi, penalizzando deviazioni (esempio: peso alla nascita nell'uomo).
- **Disruptiva**: favorisce valori estremi in entrambe le direzioni, disfavorendo valori intermedi.

Nel contesto linguistico, la lenizione consonantica rappresenterebbe selezione direzionale verso articolazioni più deboli. La resistenza alla neutralizzazione di contrasti fonologici rappresenta selezione stabilizzante attorno a configurazioni distintive. La polarizzazione sociolinguistica in comunità divise potrebbe rappresentare selezione disruptiva.

Matematicamente, questi corrispondono a diverse forme del potenziale $V(x)$:

- **Direzionale**: $V(x) = -ax$ (lineare, favorisce una direzione)
- **Stabilizzante**: $V(x) = k(x - x_0)^2$ (parabolico, favorisce un punto)
- **Disruptiva**: $V(x) = -k(x - x_0)^2 + cx^4$ (doppio pozzo, favorisce estremi)

La dinamica risultante mostra comportamenti qualitativamente diversi: convergenza monotona, oscillazioni smorzate, o bifurcazioni verso stati multipli.

11. Hard Sweeps vs Soft Sweeps: Fissazione di Innovazioni Vantaggiose

Quando una mutazione vantaggiosa appare e si fissa in una popolazione, può seguire due scenari:

- **Hard Sweep**: La mutazione appare una sola volta e sale rapidamente a frequenza alta, "spazzando via" variazione in siti genetici linkati (hitch-hiking).
- **Soft Sweep**: La mutazione appare multiple volte indipendentemente, o era già presente a bassa frequenza quando divenne vantaggiosa, cosicché multiple copie indipendenti contribuiscono alla fissazione.

I due scenari lasciano firme genetiche distinguibili: gli hard sweeps creano lunghe regioni di bassa diversità (valli di eterozigosità), mentre i soft sweeps mantengono più diversità.

Nel contesto linguistico, consideriamo l'adozione di un'innovazione fonologica vantaggiosa. Un hard sweep corrisponderebbe a un'innovazione che origina in un singolo individuo e si diffonde per imitazione sociale, portando con sé "varianti linkate" (altri aspetti idiosincratici del socioletto di quell'individuo). Un soft sweep corrisponderebbe a un'innovazione scoperta indipendentemente da multipli individui, o presente come variante minoritaria che diventa improvvisamente prestigiosa.

12. Deriva Genetica e Dimensione Effettiva della Popolazione

Un concetto centrale in genetica delle popolazioni è N_e, la dimensione effettiva della popolazione, che può differire sostanzialmente dalla dimensione censuaria N. Fattori che riducono N_e includono:

- **Sex ratio sbilanciato**: N_e = 4N_m N_f / (N_m + N_f), massimizzato quando N_m = N_f
- **Varianza nella prole**: individui con molti discendenti riducono N_e
- **Fluttuazioni demografiche**: colli di bottiglia riducono drasticamente N_e

In linguistica, la dimensione effettiva della popolazione linguistica dipende da fattori analoghi:

- **Distribuzione dello speaking time**: se pochi individui dominano la conversazione, N_e è ridotto
- **Reti sociali**: strutture centralizzate riducono N_e rispetto a reti più egalitarie
- **Alfabetizzazione**: l'ortografia standardizzata può "congelare" varianti, riducendo l'effetto della variazione parlata

Stimare N_e linguistico è difficile ma cruciale per interpretare tassi di cambiamento osservati. Una popolazione linguistica di milioni potrebbe avere N_e effettivo di solo migliaia se la maggior parte della trasmissione linguistica è mediata da pochi parlanti influenti (media, istruzione formale).

13. Modelli Multi-Locus: Ricostruzioni Filogenetiche

In genetica, l'uso di dati multi-locus permette ricostruzioni filogenetiche robuste. Diversi loci hanno storie evolutive indipendenti (sotto neutralità e linkage equilibrium), e la concordanza tra alberi genici fornisce evidenza per l'albero delle specie.

Il coalescente multi-locus gestisce le discordanze tra alberi genici dovute a incomplete lineage sorting (ILS): quando speciazioni avvengono rapidamente, polimorfismi ancestrali possono persistere attraverso eventi di speciazione, creando discordanza tra albero genico e albero delle specie.

La probabilità che un gene scelto casualmente mostri topologia discordante dall'albero delle specie dipende dal rapporto tra tempo tra speciazioni T_spec e tempo di coalescenza 2N_e. Per T_spec >> 2N_e, concordanza è alta; per T_spec << 2N_e, concordanza può essere bassa come 1/3.

In linguistica comparativa, usiamo multipli "caratteri" (cognati lessicali, innovazioni fonologiche condivise, tratti morfologici) per ricostruire alberi. Le discordanze tra caratteri possono riflettere:

- **Prestito orizzontale**: analogo al gene flow
- **Convergenza**: evoluzione parallela indipendente
- **Retention differenziale**: analogo a ILS

Metodi filogenetici Bayesiani permettono di inferire simultaneamente l'albero e tassi di cambiamento per ciascun carattere, marginalizzando su incertezze.

14. Selezione Purificante e Vincoli Funzionali

Gran parte del genoma è sotto selezione purificante (negativa), che rimuove mutazioni deleterie. Questo crea vincoli evolutivi: regioni funzionalmente importanti evolvono lentamente poiché la maggior parte delle mutazioni lì è dannosa.

In linguistica, vincoli funzionali analoghi operano: contrasti fonologici cruciali per mantenere distinzioni lessicali sono preservati, mentre contrasti ridondanti possono essere persi. La morfologia essenziale per la chiarezza sintattica è mantenuta, mentre marcature ridondanti tendono a erodersi.

Possiamo quantificare l'intensità della selezione purificante confrontando tassi di cambiamento in elementi funzionali vs non-funzionali. Per esempio, i morfemi grammaticali cambiano più lentamente del vocabolario referenziale, analogamente a come regioni codificanti proteiche evolvono più lentamente di regioni non-codificanti.

Il rapporto dN/dS (non-sinonime/sinonime) usato in genetica molecolare ha un analogo linguistico nel rapporto tra tassi di cambiamento che alterano significato vs cambiamenti fonologici puramente allofoni.

15. Speciazione Allopatrica vs Simpatrica

In biologia evoluzionistica, si distingue tra:

- **Speciazione allopatrica**: popolazioni geograficamente isolate divergono fino a incompatibilità riproduttiva
- **Speciazione simpatrica**: divergenza all'interno di una regione geografica, spesso via selezione disruptiva

La speciazione linguistica è prevalentemente allopatrica: dialetti separati geograficamente accumulano differenze fino a mutua incomprensibilità. Tuttavia, esistono casi che potrebbero essere considerati "simpatrici": emergenza di vernacoli urbani distinti in città multilingui, o divergenza di registri sociolinguistici in stratificazione verticale.

La modellizzazione richiede di specificare meccanismi che mantengono barriere riproduttive (o comunicative) in presenza di opportunità di scambio. In biologia, questo può essere assortative mating; in linguistica, potrebbe essere segregazione sociale, diglossia, o selezione di interlocutori basata su identità.

16. Riconciliando Neutralismo e Selezionismo: Sintesi Moderna

Il dibattito tra neutralisti e selezionisti in genetica si è evoluto verso una sintesi: molto polimorfismo è effettivamente neutro o nearly neutral, ma la selezione purificante è pervasiva, e la selezione positiva occasionalmente guida cambiamenti adattativi significativi.

Similmente in linguistica, possiamo riconoscere che:

- Molto della variazione fonetica fine è probabilmente neutra o nearly neutral
- Vincoli funzionali (distintività, learnability) impongono selezione purificante
- Pressioni socioeconomiche (prestigio, identità) guidano occasionalmente cambiamenti direzionali rapidi

I modelli SDE permettono di incorporare tutti questi aspetti simultaneamente, con drift neutrale come baseline e selezione come perturbazione dipendente dal contesto. L'inferenza statistica può poi determinare l'importanza relativa di ciascuna forza per fenomeni specifici.

Metodi Computazionali, Inferenza Statistica e Prospettive Future

1. Simulazione Numerica di Equazioni Differenziali Stocastiche

L'analisi di modelli SDE complessi richiede quasi sempre simulazione numerica, poiché soluzioni analitiche esatte sono disponibili solo per casi speciali. Lo schema di Eulero-Maruyama, già menzionato, rappresenta il metodo base ma esistono schemi più sofisticati con migliori proprietà di convergenza.

Lo schema di Milstein, per esempio, incorpora termini correttivi del secondo ordine derivati dal lemma di Itô:

$$X_{n+1} = X_n + \mu(X_n)\,\Delta t + \sigma(X_n)\,\Delta W_n + \frac{1}{2}\,\sigma(X_n)\,\sigma'(X_n)\,[(\Delta W_n)^2 - \Delta t]$$

dove σ' è la derivata di σ rispetto a X, e $\Delta W_n \sim N(0, \Delta t)$. Questo schema ha convergenza di ordine 1 (rispetto al passo temporale) per il strong error, rispetto all'ordine 0,5 di Eulero-Maruyama.

Per sistemi ad alta dimensionalità o con drift e diffusione complessi, schemi ancora più sofisticati sono necessari. I metodi Runge-Kutta stocastici (SRK) generalizzano i classici metodi RK deterministici. Implementazioni efficienti richiedono attenzione alla generazione di numeri casuali, alla gestione della memoria per lunghe traiettorie, e alla parallelizzazione quando si simulano ensembles.

Nel contesto linguistico, simulare l'evoluzione di un intero sistema fonologico su scale temporali di secoli richiede:

1. **Discretizzazione temporale appropriata**: passi troppo grandi introducono errori, passi troppo piccoli sono computazionalmente costosi. Tipicamente $\Delta t \sim 0{,}1\text{-}1$ generazioni.

2. **Rappresentazione dello spazio fonetico**: griglie discrete vs rappresentazioni continue, dimensionalità ridotta vs completa.

3. **Condizioni al contorno**: riflessione o assorbimento ai bordi dello spazio fonetico fisicamente accessibile.

4. **Accoppiamenti**: gestione efficiente delle interazioni tra fonemi per sistemi con molti fonemi.

2. Inferenza Bayesiana per Parametri di SDE

Dato un dataset di osservazioni linguistiche (per esempio, realizzazioni fonetiche campionate a diversi punti temporali), vogliamo inferire i parametri sottostanti del modello SDE: coefficienti di drift μ, coefficienti di diffusione σ, e eventuali iper-parametri come dimensioni di popolazione N_e.

L'approccio bayesiano specifica distribuzioni prior sui parametri e calcola la distribuzione posteriore dato i dati tramite il teorema di Bayes:

$p(\theta \mid \text{dati}) \propto p(\text{dati} \mid \theta)\, p(\theta)$

Il termine chiave è la likelihood p(dati | θ), che per SDE non ammette forma chiusa in generale. Dobbiamo quindi usare metodi di approssimazione.

Un approccio è la likelihood approssimata basata su transizioni discrete. Se osserviamo $X(t_1)$, $X(t_2)$, ..., $X(t\square)$, possiamo approssimare:

$p(X_{i+1} \mid X_i, \theta) \approx N(X_i + \mu(X_i, \theta)\Delta t, \Sigma(X_i, \theta)\Delta t)$

dove $\Delta t = t_{i+1} - t_i$ e $\Sigma = \sigma\sigma^T$. Questa approssimazione è esatta per SDE lineari, approssimata per SDE non-lineari.

Un'alternativa più sofisticata usa il filtro di Kalman esteso o il particle filter per calcolare la likelihood esatta per i dati parzialmente osservati, propagando distribuzioni attraverso le dinamiche stocastiche.

3. Metodi MCMC e Approximate Bayesian Computation

Con likelihood complesse o intrattabili, i metodi Markov Chain Monte Carlo (MCMC) permettono di campionare dalla distribuzione posteriore senza calcolarla esplicitamente. L'algoritmo di Metropolis-Hastings itera:

1. Proponi nuovo parametro θ* dalla distribuzione proposta q(θ* | θ_current)
2. Calcola acceptance ratio $\alpha = [p(\text{dati} \mid \theta^*)\, p(\theta^*)\, q(\theta_{current} \mid \theta^*)] / [p(\text{dati} \mid \theta_{current})\, p(\theta_{current})\, q(\theta^* \mid \theta_{current})]$
3. Accetta θ* con probabilità min(1, α), altrimenti mantieni θ_current

Dopo un burn-in iniziale, i campioni formano una catena di Markov la cui distribuzione stazionaria è la posteriore desiderata.

Per SDE dove la likelihood è completamente intrattabile, Approximate Bayesian Computation (ABC) offre un'alternativa. L'idea è simulare dati sintetici dal modello con parametri proposti θ*, e accettare θ* se i dati simulati sono "sufficientemente simili" ai dati osservati secondo qualche metrica di distanza d(·,·):

Accetta θ* se d(dati_sim, dati_obs) < ε

dove ε è una tolleranza. Questo evita di calcolare la likelihood esplicitamente, al costo di introdurre errore controllato da ε.

Nel contesto linguistico, potremmo usare ABC per inferire parametri di modelli SDE per il Great Vowel Shift, simulando traiettorie fonetiche e confrontandole con ricostruzioni storiche.

4. Maximum Likelihood Estimation via Transitional Densities

Per SDE con struttura speciale, sono disponibili metodi di maximum likelihood più diretti. Se la SDE ammette soluzioni in forma chiusa per le densità di transizione p(X_{t+s} | X_t), possiamo costruire la log-likelihood:

$\ell(\theta) = \Sigma_i \log p(X_{t_{i+1}} \mid X_{t_i}, \theta)$

e massimizzarla numericamente rispetto a θ usando algoritmi di ottimizzazione come Newton-Raphson o BFGS.

Anche quando le densità di transizione non sono esplicite, approssimazioni numeriche possono essere ottenute risolvendo l'equazione di Fokker-Planck in avanti, o usando espansioni perturbative.

Un metodo sofisticato è la likelihood-free inference basata su metodi di score matching, che evitano di normalizzare la densità stimando direttamente lo score (gradiente del log-likelihood).

5. Model Selection e Confronto di Ipotesi

Dati multipli modelli candidati M_1, M_2, ..., M□ per lo stesso fenomeno linguistico, vogliamo selezionare quello che bilancia meglio fit ai dati e complessità del modello. Il framework bayesiano offre strumenti naturali attraverso il Bayes factor:

$BF_{12} = p(\text{dati} \mid M_1) / p(\text{dati} \mid M_2)$

dove $p(\text{dati} \mid M_i)$ è la marginal likelihood (evidenza) per il modello i:

$p(\text{dati} \mid M_i) = \int p(\text{dati} \mid \theta, M_i)\, p(\theta \mid M_i)\, d\theta$

Un $BF_{12} > 1$ favorisce M_1, mentre $BF_{12} < 1$ favorisce M_2. Interpretazioni convenzionali considerano BF > 10 come evidenza forte, BF > 100 come evidenza decisiva.

Calcolare la marginal likelihood richiede integrare su tutto lo spazio dei parametri, computazionalmente sfidante. Metodi come nested sampling, bridge sampling, o thermodynamic integration sono stati sviluppati per questo scopo.

Alternative frequentiste includono criteri di informazione come AIC (Akaike Information Criterion):

$$AIC = -2 \log \hat{L} + 2k$$

o BIC (Bayesian Information Criterion):

$$BIC = -2 \log \hat{L} + k \log n$$

dove $\hat{L}$ è la likelihood massimizzata, k il numero di parametri, e n la dimensione del campione. Modelli con AIC/BIC più bassi sono preferiti.

6. Cross-Validation e Predizione Out-of-Sample

Un test robusto di capacità predittiva divide i dati in training set e test set. Il modello è stimato sul training set e valutato sul test set. K-fold cross-validation ripete questo processo K volte, ciascuna con un diverso fold tenuto fuori.

Per dati temporali, è appropriato time-series cross-validation dove il training set consiste sempre in osservazioni prima del test set (rispettando la causalità temporale).

Nel contesto linguistico, potremmo addestrare un modello SDE sui primi tre secoli della GVS e predire le configurazioni fonetiche nel quarto secolo, confrontando con ricostruzioni storiche indipendenti. Il prediction error quantifica il potere predittivo del modello.

7. Dimensionality Reduction e Manifold Learning

Gli spazi fonetici reali, sebbene formalmente ad alta dimensionalità (ogni parametro articolatorio aggiunge una dimensione), spesso giacciono vicino a varietà (manifolds) di dimensione intrinseca molto più bassa. Questo riflette vincoli articolatori e percettivi che correlano diverse dimensioni.

Tecniche di manifold learning come Isomap, Locally Linear Embedding (LLE), o diffusion maps possono estrarre rappresentazioni di dimensione ridotta che preservano struttura locale. Applicare queste tecniche a dati fonetici può rivelare la "vera" dimensionalità effettiva dello spazio fonetico.

Una volta identificata la varietà intrinseca, possiamo formulare SDE direttamente su questa varietà, riducendo drammaticamente la complessità computazionale. Le SDE su varietà riemanniane sono un'area matematicamente sofisticata ma con crescente applicazione pratica.

8. Deep Learning e Neural SDEs

Un sviluppo recente combina equazioni differenziali stocastiche con deep learning. Neural SDEs parametrizzano i coefficienti di drift e diffusione con reti neurali:

$dX(t) = f_\theta(X(t), t)\, dt + g_\varphi(X(t), t)\, dW(t)$

dove f_θ e g_φ sono reti neurali con parametri θ e φ. Questi parametri sono appresi end-to-end da dati temporali usando tecniche di backpropagation attraverso la SDE.

I vantaggi includono:

- **Flessibilità**: reti neurali possono approssimare funzioni arbitrariamente complesse
- **Feature learning**: la rete impara automaticamente rappresentazioni rilevanti
- **Scalabilità**: moderne tecniche di ottimizzazione gestiscono milioni di parametri

Applicazioni in linguistica potrebbero includere modellizzare traiettorie di cambiamento fonologico senza specificare forme funzionali a priori, lasciando che la rete scopra le regolarità dai dati.

9. Agent-Based Models e Simulazioni Multi-Scala

Un approccio complementare alla modellizzazione continua è la simulazione agent-based (ABM), dove ciascun parlante individuale è rappresentato esplicitamente come agente autonomo che interagisce con altri secondo regole locali.

Gli ABM permettono di incorporare eterogeneità individuale, struttura di rete sociale, apprendimento, e altri dettagli microscopici difficili da catturare in modelli continui. La dinamica emergente dell'ensemble può poi essere confrontata con le predizioni dei modelli SDE.

Un workflow integrato potrebbe:

1. Simulare ABM dettagliato con migliaia di agenti
2. Estrarre dinamiche aggregata (medie e varianze temporali)
3. Fittare modelli SDE ridotti alle dinamiche aggregate
4. Validare che il modello SDE catturi le caratteristiche essenziali

Questo approccio multi-scala connette meccanismi microscopici a fenomeni macroscopici, fornendo intuizioni su quali dettagli sono cruciali e quali possono essere coarse-grained.

10. Analisi di Sensitività e Incertezza

I modelli SDE per fenomeni linguistici contengono inevitabilmente numerosi parametri con incertezza. L'analisi di sensitività quantifica come l'incertezza negli input (parametri, condizioni iniziali) si propaga agli output (predizioni).

Metodi di analisi di sensitività globale come gli indici di Sobol decompongono la varianza dell'output in contributi da ciascun parametro e loro interazioni:

$\mathrm{Var}(Y) = \Sigma_i V_i + \Sigma_{i<} V_{i} + \ldots$

dove $V_i = Var[E(Y \mid X_i)]$ è l'effetto del primo ordine del parametro i, e V_i□ cattura interazioni. Gli indici di sensitività normalizzati $S_i = V_i / Var(Y)$ sommano a 1.

Identificare i parametri con S_i alto permette di focalizzare sforzi di misura e calibrazione su quelli che più influenzano le predizioni, mentre parametri con S_i piccolo possono essere fissati senza grande perdita.

11. Validazione con Dati Storici e Filologici

La validazione ultima dei modelli SDE linguistici richiede confronto con documentazione storica reale. Per lingue ben documentate come l'inglese, il francese, o il cinese, abbiamo testimonianze scritte che coprono secoli o millenni, permettendo confronti quantitativi.

Le sfide includono:

- **Incompletezza**: documentazione storica è frammentaria e sociolinguisticamente stratificata
- **Bias ortografico**: sistemi di scrittura conservatori mascherano cambiamenti fonetici
- **Interpretazione**: ricostruzioni fonetiche da ortografie antiche richiedono inferenza

Metodi bayesiani possono incorporare queste incertezze, trattando le ricostruzioni fonetiche stesse come variabili latenti da inferire congiuntamente con i parametri del modello.

12. Predizioni Controfattuali e Esperimenti In Silico

Una volta validato, un modello SDE permette esperimenti controfattuali impossibili nella realtà: cosa sarebbe successo se la popolazione fosse stata 10 volte più grande? Se il contatto con altre lingue fosse stato assente? Se certe pressioni selettive fossero state invertite?

Questi esperimenti in silico forniscono intuizioni causali che vanno oltre correlazioni osservazionali. Per esempio, potremmo testare l'ipotesi che il GVS fu innescato dall'influsso normanno simulando scenari con e senza massicci prestiti francesi.

Ensemble di simulazioni controfattuali permettono anche di quantificare quanto "inevitabile" fosse un certo cambiamento linguistico: se praticamente tutte le simulazioni producono risultati simili, il cambiamento era robusto; se i risultati variano ampiamente, il cambiamento era contingente.

13. Applicazioni alla Documentazione di Lingue in Pericolo

Le tecniche di modellizzazione SDE hanno potenziali applicazioni pratiche nella linguistica documentale e nella conservazione linguistica. Lingue in pericolo critico spesso mostrano variabilità aumentata e erosione rapida di contrasti, fenomeni modellizzabili come aumento della diffusione e drift verso configurazioni semplificate.

Modelli predittivi potrebbero:

- Stimare tassi di cambiamento e estinzione proiettati

- Identificare caratteristiche a maggior rischio di perdita
- Informare strategie di rivitalizzazione focalizzando su elementi cruciali
- Monitorare salute linguistica attraverso metriche quantitative (analoghe a indicatori ecologici)

Questo richiede sensibilità etica: le comunità linguistiche devono essere centrali nel processo, e i modelli matematici devono informare ma non dettare decisioni.

14. Integrazione con Neuroscienze Cognitive

Le basi cognitive e neurali del cambiamento linguistico sono un'area emergente. Studi di neuroimaging mostrano come il cervello elabora e produce linguaggio, e come questi processi variano inter-individualmente e cambiano con l'età. Possiamo cercare di ancorare i parametri dei modelli SDE a meccanismi neurali misurabili. Per esempio:

- Il coefficiente di diffusione σ potrebbe relazionarsi alla variabilità neurale nella rappresentazione fonetica
- Tassi di apprendimento potrebbero connettere a plasticità sinaptica e critical periods
- Effetti di prestigio potrebbero relazionarsi a circuiti di reward e motivazione sociale

Questa grounding neurale trasformerebbe i modelli SDE da descrizioni fenomenologiche a modelli meccanicistici con potere esplicativo più profondo.

15. Verso una Teoria Unificata: Sintesi e Prospettive Future

L'applicazione di equazioni differenziali stocastiche alla linguistica storica rappresenta un passo verso una teoria quantitativa unificata dell'evoluzione linguistica. Come la sintesi moderna ha unificato genetica mendeliana ed evoluzione darwiniana, una sintesi linguistica potrebbe unificare linguistica storica, sociolinguistica variazionista, e linguistica cognitiva in un framework matematico comune.

Gli ingredienti chiave di tale sintesi includerebbero:

1. **Modelli multi-scala**: dal neurone individuale alle comunità linguistiche
2. **Dinamiche stocastiche**: incorporando sia drift neutrale sia selezione funzionale
3. **Struttura di rete**: modellizzando esplicitamente chi parla con chi
4. **Apprendimento**: come nuove generazioni acquisiscono e innovano
5. **Contatto**: flussi orizzontali tra linee verticali di trasmissione

Diverse direzioni di ricerca futura appaiono promettenti:

A. Modelli più Realistici di Trasmissione Andare oltre assunzioni semplicistiche di trasmissione verticale perfetta, incorporando:

- Acquisizione da input multi-parlante distribuito
- Apprendimento imperfetto e rianalisi generazionali
- Code-switching e translanguaging in contesti multilingui

B. Incorporazione di Vincoli Cognitivi Informare i potenziali $V(X)$ con dati sperimentali su:

- Facilità articolatoria misurata biomeccanicamente
- Distintività percettiva misurata psicoacusticamente
- Complessità rappresentazionale misurata neuralmente

C. Analisi di Big Data Linguistici Sfruttare corpus massivi (social media, depositi digitali storici):

- Tracking di cambiamenti in tempo reale
- Risoluzione geografica e sociale fine
- Identificazione di innovatori e early adopters

D. Causazione Inversa e Feedback Modellizzare come cambiamenti linguistici retroagiscono sulla struttura sociale:

- Divergenza linguistica che crea barriere sociali
- Convergenza che facilita integrazione
- Co-evoluzione lingua-società

E. Estensione a Morfosintassi La maggior parte del lavoro si è focalizzata su fonologia, ma:

- Cambiamento morfologico (semplificazione, analogia)
- Grammaticalizzazione di lessico in morfologia
- Deriva sintattica (ordine di parole, allineamento)

richiedono rappresentazioni degli spazi di stati più astratte ma le stesse tecniche stocastiche dovrebbero applicarsi.

16. Riflessioni Epistemologiche: Modelli e Realtà

È importante mantenere umiltà epistemologica riguardo allo status dei modelli matematici in linguistica. George Box scrisse che "tutti i modelli sono sbagliati, ma alcuni sono utili". I modelli SDE non catturano la "realtà vera" del cambiamento linguistico in tutta la sua ricchezza, ma forniscono lenti utili per organizzare pensiero e dati.

I benefici dei modelli formali includono:

- **Precisione**: affermazioni vaghe diventano predizioni quantitative testabili
- **Coerenza**: assunzioni devono essere esplicitate e loro conseguenze derivate rigorosamente
- **Generalità**: insights da un dominio (genetica) informano altri (linguistica)
- **Falsificabilità**: modelli possono essere rifiutati se predizioni falliscono

I rischi includono:

- **Over-fitting**: modelli complessi possono fittare rumore piuttosto che segnale
- **Reificazione**: dimenticare che categorie del modello sono costrutti, non realtà
- **Falsa precisione**: numeri spuriamente precisi mascherando incertezze profonde

Il percorso saggio è pluralismo metodologico: usare modelli formali come uno strumento tra molti, in dialogo con filologia, fieldwork, e teoria qualitativa.

17. Conclusione: Un Programma di Ricerca Aperto

L'applicazione di equazioni differenziali stocastiche alla deriva linguistica e al cambiamento fonologico rappresenta un programma di ricerca giovane e vibrante. Abbiamo visto come i concetti fondamentali della teoria stocastica - moto browniano, drift, diffusione, salti, processi di nascita e morte, coalescenza - si mappano naturalmente su fenomeni linguistici. Le analogie profonde con la genetica delle popolazioni non sono superficiali ma riflettono parallelismi strutturali autentici tra sistemi che evolvono attraverso trasmissione con variazione. Sia geni che lingue sono sistemi informativi complessi soggetti a forze sia stocastiche sia selettive, operanti su multiple scale temporali e spaziali.

I successi finora includono:

- Modellizzazione quantitativa della Grande Mutazione Vocalica e fenomeni simili
- Ricostruzione filogenetica robusta di famiglie linguistiche
- Quantificazione di tassi di cambiamento e tempi di divergenza
- Predizione di pattern di diversità linguistica da primi principi

Le sfide rimangono:

- Dati storici frammentari e rumorosi
- Alta dimensionalità degli spazi linguistici
- Complessità delle interazioni sociolinguistiche
- Difficoltà di validazione per eventi lontani nel passato

Tuttavia, la convergenza di migliori dati (digitalizzazione massiva, monitoraggio real-time via social media), migliori metodi computazionali (simulazioni massive, machine learning, inferenza bayesiana scalabile), e migliori fondamenti teorici (teoria stocastica rigorosa, connessioni con neuroscienze) promette progressi significativi nei decenni venturi. Il sogno ultimo sarebbe una teoria predittiva quantitativa dell'evoluzione linguistica, analoga in scope alla sintesi moderna biologica. Sebbene questo obiettivo rimanga distante, i passi compiuti mostrano che non è utopico. La linguistica storica, da disciplina prevalentemente qualitativa e umanistica, sta diventando anche una scienza quantitativa e predittiva, capace di dialogo proficuo con altre scienze dei sistemi complessi. Questo arricchimento metodologico non impoverisce la ricchezza interpretativa della filologia tradizionale, ma la complementa. I modelli matematici pongono nuove domande, richiedono nuovi tipi di dati, e offrono nuove intuizioni - ma il significato di questi risultati deve sempre essere interpretato nel contesto più ampio della storia, cultura e cognizione umana. In definitiva, studiare il cambiamento linguistico attraverso equazioni differenziali stocastiche ci insegna qualcosa di profondo sulla natura del linguaggio stesso: che è un sistema complesso adattativo, né completamente deterministico né completamente caotico, ma caratterizzato da una delicata interazione tra regolarità strutturale e contingenza storica, tra necessità funzionale e accidente evolutivo. E forse, in questo equilibrio dinamico, risiede parte della risposta al mistero di come il linguaggio rimanga allo stesso tempo stabile abbastanza da permettere comunicazione

attraverso generazioni, e flessibile abbastanza da adattarsi continuamente alle mutevoli esigenze dei suoi utenti.

Bibliografia Essenziale

Matematica Stocastica:

- Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Springer.
- Karatzas, I., & Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*. Springer.
- Pavliotis, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer.

Genetica delle Popolazioni:

- Ewens, W. J. (2004). *Mathematical Population Genetics*. Springer.
- Gillespie, J. H. (2004). *Population Genetics: A Concise Guide*. Johns Hopkins University Press.
- Wakeley, J. (2008). *Coalescent Theory: An Introduction*. Roberts and Company.

Linguistica Storica e Quantitativa:

- Labov, W. (2001). *Principles of Linguistic Change: Social Factors*. Blackwell.
- McMahon, A., & McMahon, R. (2012). *Evolutionary Linguistics*. Cambridge University Press.
- Dunn, M. (Ed.). (2015). *The Oxford Handbook of Linguistic Typology*. Oxford University Press.

Filogenetica Linguistica:

- Dunn, M. (2014). Language phylogenies. In *The Routledge Handbook of Historical Linguistics*.
- Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426, 435-439.
- Greenhill, S. J., & Gray, R. D. (2009). Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In *Discovering History Through Language: Papers in Honour of Malcolm Ross*.

Applicazioni SDE in Linguistica:

- Blythe, R. A., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88(2), 269-304.
- Baxter, G. J., et al. (2006). Utterance selection model of language change. *Physical Review E*, 73(4), 046118.
- Newberry, M. G., et al. (2017). Detecting evolutionary forces in language change. *Nature*, 551, 223-226.